N° d'ordre 254

UNIVERSITE BLAISE PASCAL - CLERMONT-FERRAND

UN SYSTEME DE PERCEPTION ATTENTIONNELLE POUR LES ROBOTS MOBILES EN ENVIRONNEMENT EXTERIEUR

Mémoire présenté par

Frédéric CHAUSSE

Maître de Conférences

à l'Institut Universitaire de Technologie
de l'Université d'Auvergne

en vue de l'obtention de

l'Habilitation à Diriger des Recherches

Soutenu publiquement le 23 mai 2008 devant le jury composé de :

Claude	BARROUIL	Directeur de Recherche à l'ONERA	Examinateur
Roland	CHAPUIS	Professeur à Polytech' Clermont	Rapporteur UBP
Michel	DEVY	Directeur de Recherche au LAAS/CNRS	Rapporteur
Marc	RICHETIN	Professeur à l'Université Blaise Pascal	Examinateur et Président du jury
Patrick	RIVES	Directeur de Recherche à l'INRIA	Rapporteur
Mme Claudine	SCHMIDT-LAINE	Directrice de Recherche au CNRS	Examinatrice
John K.	TSOTSOS	Professeur à l'Université de York (Canada)	Rapporteur

[1] M.-C.. CHAUSSE et J.-P. CHAUSSE, *Frédéric*, Clermont-Ferrand, Juin 1968.

[2] S. CHAUSSE et F. CHAUSSE, *Juliette*, Clermont-Ferrand, Juillet 1999.

[3] S. CHAUSSE et F. CHAUSSE, *Jeanne*, Clermont-Ferrand, Novembre 2002.

Remerciements

Je commencerai par remercier Marc Richetin, Jean-Paul Germain et Michel Dhome qui se sont succédés à la direction du LASMEA position de laquelle ils ont gardé un œuil bienveillant sur mon travail et qui œuvrent sans relâche pour que les chercheurs du LASMEA puissent bénéficier des meilleures conditions de travail.

MM Michel Devy, Patrick Rives et John, K. Tsotsos ont assuré la lourde tâche de rapporteur, acceptant de venir de loin et de modifier leur emploi du temps pour évaluer mon travail. Je les remercie également pour cela et pour les remarques constructives qu'ils ont émises.

Que Mme Claudine Schmidt-Lainé, MM Claude Barrouil et Marc Richetin, qui ont examiné mon travail et participé au jury trouvent ici l'expression de ma reconnaissance. J'ajoute une mention spéciale à Marc Richetin qui a accepté de le présider.

Mon travail aurait une valeur bien moindre sans l'investissement au quotidien de mes collègues du LASMEA permanents ou temporaires en particulier MM Serge Alizon, Alain Blanc, Sébastien Bonnet, Laurent Malaterre et François Marmoiton qui assurent dans des conditions parfois difficiles le bon fonctionnement des plates-formes robotiques grâces auxquelles la validation expérimentale des méthodes est possible. Merci à eux !

Dans un autre registre, je dois également remercier Eliane De Dea et avec elle l'ensemble du service administratif du LASMEA qui font l'impossible afin d'assurer le bon déroulement de nos missions. Leur place est une inconfortable interface entre nous et les voies impénétrables de la bureaucratie universitaire.

En respectant l'ordre chronologique, je remercie Romuald, Jean, Cédric, Noël, Nadir, Laurence et David, les anciens et actuels doctorants avec qui j'ai travaillé. Il faut leur rendre le mérite d'une bonne partie des résultats présentés dans ce mémoire qui ne tente peut-être que de mettre leur travail en valeur.

J'ai pu garder depuis quelques années des contacts fréquents avec le Cemagref et en particulier avec Michel Berducat et Christophe Debain. Ils ont su créer une ambiance de travail particulièrement agréable qui n'est pas étrangère à la grande qualité des résultats scientifiques que nous avons obtenus ensembles. Je les en remercie chaleureusement.

Je garde une pensée reconnaissante pour les "anciens", Patrick Ackerman, Joseph Alizon, Pierre Andanson, Jean-Pierre Battut, Jean-Pierre Chausse, Jean Gallice, Jean-Pierre Imbaud, Jean-Louis Laporte, Michel Naranjo, Christian Thibaud, qui ont tenté avec courage de m'éclairer des lumières de leur expérience. Ils m'ont mis le pieds à certains étriers et rendu ainsi bien plus confortable le début de ma carrière.

Vouloir dresser la liste de ce que je dois à Roland Chapuis est illusoire. C'est lui qui a pensé initialement le principe du système de perception sur lequel repose mon travail. C'est lui qui a contribué à mon intégration au LASMEA notamment en m'associant systématiquement à l'encadrement des thèses dont il avait la direction. Sur un plan plus personnel, nous partageons entre autre le goût des discussions rallongées sur tout un tas de sujets divers et variés, un attachement profond à la Terre de Haute-Loire, l'attrait pour certaines musiques bruyantes où les guitares saturent volontairement les amplis. . .Je m'arrête car là encore la série est trop longue. Alors pour tout cela et pour le reste, pour tout ce que j'oublie, pour tout ce qui arrivera encore je l'espère, Roland, je te dis simplement merci.

Table des matières

I Notice individuelle **11**

1 Etat civil - Diplômes - Emplois **13**

1.1 Etat civil et situation de famille 13

1.2 Diplômes et titres universitaires 13

1.3 Emplois et fonctions 14

2 Activités d'enseignement et d'intérêt collectif **15**

2.1 Service statutaire 15

2.2 Responsabilités et enseignements au C.N.A.M. 15

2.3 Formation continue 16

2.4 Tâches d'intérêt collectif 17

3 Activités de recherche **19**

3.1 Contexte 19

3.2 Encadrement scientifique 19

3.2.1 Post-doctorants 19

3.2.2 Doctorants 19

3.2.3 Stagiaires D.E.A. 20

3.3 Publications et communications 20

3.3.1 Thèse 20

3.3.2 Articles dans des revues nationales et internationales à comité de lecture 21

3.3.3 Communications à des conférences internationales avec actes et comité de lecture 21

3.3.4 Communications écrites à des conférences nationales avec actes et comité de lecture 23

3.3.5 Colloques sans comité de sélection 24
3.4 Collaborations de recherche 24
3.5 Groupes de recherche nationaux 25

II Travaux de recherche 27

4 Systèmes de perception 31
4.1 Introduction 31
4.1.1 Le modèle de Marr : la vision par ordinateur 31
4.1.2 Le modèle d'Aloimonos : la vision active 32
4.1.3 Le modèle de Bajcsy : la perception active 33
4.2 Perception attentionnelle 33
4.2.1 Bilan 35
4.3 Modèle de perception attentionnelle proposé 35
4.3.1 Représentation synthétique de la scène 36
4.3.2 Apprentissage 37
4.3.3 Evolution du processus attentionnel : la focalisation progressive 37
4.3.4 Critère d'arrêt : évaluation de la vraisemblance et de la précision 39
4.4 Conclusion 39

5 Perception pour la localisation 41
5.1 Localisation d'un véhicule sur sa voie de circulation 41
5.1.1 Introduction 41
5.2 Description détaillée de l'application 42
5.2.1 Motivations de l'application et état de l'art en détection de bord de voies . 42
5.2.2 Reconnaissance 44
5.2.3 Modèle 44
5.2.4 Localisation et suivi 48
5.2.5 Gestion du nombre de voies 49
5.2.6 Résultats 50
5.3 Localisation d'un véhicule dans un environnement cartographié 52
5.3.1 Introduction 52

5.3.2 Description détaillée de l'application 52
5.3.3 Localisation par couplage vision/carte 57
5.3.4 Résultat de localisation 60
5.4 Localisation et guidage d'un robot mobile autonome 64
5.4.1 Introduction 64
5.4.2 Intérêt de l'application 64
5.4.3 Principe de la méthode de localisation utilisant la focalisation progressive 67
5.4.4 Traitement probabiliste des mauvaises associations de données 69
5.4.5 Expérimentation et résultats 79
5.4.6 Conclusion et perspectives 85

6 Perception des objets dynamiques 87
6.1 Introduction 87
6.2 Reconnaissance des objets par vision focalisée 88
6.2.1 Introduction 88
6.2.2 Focalisation de l'attention en reconnaissance d'objets 88
6.2.3 Vision focalisée 89
6.2.4 Apprentissage du modèle 91
6.2.5 Stratégie de reconnaissance 95
6.2.6 Phase de décision 98
6.3 Résultats 100
6.3.1 Résultat sur des images seules 100
6.3.2 Suivi d'objet 100
6.4 Bilan 104
6.5 Perception multisensorielle des piétons 105
6.5.1 Introduction 105
6.5.2 Description détaillée de l'application 106
6.5.3 Résultats expérimentaux 111
6.5.4 Conclusion 116

Première partie

Notice individuelle

1 Etat civil - Diplômes - Emplois

1.1 Etat civil et situation de famille

Frédéric CHAUSSE

Né le 22 Juin 1968 à Clermont-Ferrand (63), nationalité française, marié, deux enfants.

Adresse personnelle : Chemin de Naja - 63450 Saint-Saturnin

Tél. : 06-76-01-10-18

Adresses professionnelles :
– Enseignement
  I.U.T.
  Ensemble Universitaire des Cézeaux
  BP 86
  63172 AUBIERE CEDEX
  Tél. : 04-73-17-71-62
– Recherche
  L.A.S.M.E.A.
  Ensemble Universitaire des Cézeaux
  63177 AUBIERE CEDEX
  Tél. : 04-73-40-52-10

1.2 Diplômes et titres universitaires

1986 : Baccalauréat série C à Clermont-Ferrand.

1988 : D.E.U.G. Sciences des Structures et de la Matière, option E.E.A., Université Blaise Pascal, Clermont-Ferrand II, (mention Assez Bien).

1989 : First Certificate in English, University of Cambridge.

1991 : Diplôme d'ingénieur du C.U.S.T., spécialité Génie Electrique, Université Blaise Pascal, (mention Bien).

1991 : D.E.A. d'électronique et systèmes, option traitement des images et reconnaissance des formes, Université Blaise Pascal, (mention Assez Bien).

1994 : Doctorat de l'Université Blaise Pascal, Clermont-Ferrand II Spécialité Electronique (mention Très Honorable).

1.3 Emplois et fonctions

1991-1994 : Allocataire du Ministère de la Recherche et de la Technologie

1991-1994 : Moniteur d'Enseignement Supérieur, I.U.T. d'Aubière, Université d'Auvergne

1994-95 : A.T.E.R. (demi poste), IUT d'Aubière, Université d'Auvergne

1995 : Maître de Conférences stagiaire, I.U.T. d'Aubière, Université d'Auvergne

Depuis 1996 : Maître de Conférences titulaire, I.U.T. d'Aubière, Université d'Auvergne

2004-06 : Délégation de recherche au C.N.R.S.

2 Activités d'enseignement et d'intérêt collectif

2.1 Service statutaire

J'effectue l'intégralité de mes activités statutaires d'enseignement au département de Mesures Physiques de l'I.U.T de l'Université d'Auvergne. J'enseigne des cours (C), travaux dirigés (T.D.) et travaux pratiques (T.P.) qui abordent la plupart des thèmes traditionnellement rattachés à l'E.E.A. : électronique analogique et numérique, informatique industrielle, traitement du signal, automatique et électronique de puissance, propagation des signaux.

En 2004/05 et 2005/06 j'ai bénéficié d'une délégation de recherche au CNRS.

Les volumes horaires enseignés année après année sont détaillés dans le tableau 2.1. Ils totalisent 2859 Heures Equivalent T.D. (HTD) sur 13 ans soit 220HTD par an. Ceci correspond à une moyenne de 28 heures complémentaires par an.

2.2 Responsabilités et enseignements au C.N.A.M.

Je coordonne les enseignements de deux filières du Centre Régional d'Auvergne du C.N.A.M. :

Depuis 2001 : Responsable de la filière Instrumentation et Mesures option Métrologie-Qualité.

Depuis 2002 : Responsable de la filière Automatismes Industriels.

Je suis agréé par le C.N.A.M. pour plusieurs unités de valeur que j'enseigne ou ai enseignées régulièrement en cours du soir depuis 1994.

1994, 1996, 1998, 2000 : Valeur Automatismes Industriels B1, "Représentation fréquentielle appliquée à la commande des systèmes linéaires" , 24 heures de cours et 26 heures deT.D.

1995, 1996, 1998, 2000 : Valeur Automatismes Industriels B3, "Introduction aux techniques de commande des systèmes linéaires", 24 heures de cours et 26 heures deT.D.

2000, 2002, 2003, 2005 : Valeur Electronique B4, "Systèmes à microprocesseurs", 12 heures de cours, 12 heures de T.D. et 20 heures deT.P.

1999, 2001, 2002 : Valeur Instrumentation Scientifique C0, "Signal et bruits", 24 heures de cours et 26 heures deT.D.

2003 : Valeur Automatismes Industriels C5, "Identification et estimation des systèmes", 24 heures de cours et 26 heures de T.D.

J'ai encadré 4 auditeurs du C.N.A.M. pour la préparation de leur mémoire d'ingénieur (stage de 10 mois) :

Année	Discipline	Volume horaire
1995/96	Mécanique, Thermodynamique Optique, Mécanique E.E.A.	56 h de T.P. 112 h de T.P. 234 h de T.P.
1996/97	Mécanique, Thermodynamique Optique, Mécanique E.E.A. Informatique (programmation)	112 h de T.P. 112 h de T.P. 234 h de T.P. 10 h de Cours 6 h de T.D.
1997/98	Mécanique, Thermodynamique Optique, Mécanique E.E.A. Informatique (programmation)	112 h de T.P. 140 h de T.P. 234 h de T.P. 10 h de Cours 6 h de T.D.
1998/99	Mécanique, Thermodynamique Optique, Mécanique E.E.A. Informatique (programmation)	112 h de T.P. 140 h de T.P. 234 h de T.P. 10 h de Cours 6 h de T.D.
1999/00	Mécanique, Thermodynamique Optique, Mécanique E.E.A.	112 h de T.P. 140 h de T.P. 234 h de T.P.
2000/01	Optique, Mécanique E.E.A.	52 h de T.P. 234 h de T.P.
2001/02	Mécanique, Capteurs, Résist. des Matériaux E.E.A.	78 h de T.P. 234 h de T.P.
2002/03	Optique Mécanique E.E.A. Informatique Industrielle	52 h de T.P. 234 h de T.P. 9 h de Cours
2003/04	Physique Electronique des systèmes discrets E.E.A. Informatique Industrielle	4 h de T.D. 14 h de Cours 208 h de T.P. 18 h de Cours
2004/06	Délégation au C.N.R.S.	
2006/07	Informatique Industrielle Traitement du Signal Capteurs Automatique Filtrage numérique	12 h de T.D. 8 h de Cours, 16 h de T.D. et 96 h de T.P. 8 h de Cours et 14 h de T.D. 8 h de Cours, 16 h de T.D, 32 h de T.P. 10 h de Cours et 20 h de T.D.
2007/08	Informatique Industrielle Traitement du Signal Capteurs Automatique Filtrage numérique	12 h de T.D. 8 h de Cours, 16 h de T.D. et 96 h de T.P. 8 h de Cours et 14 h de T.D. 8 h de Cours, 16 h de T.D, 32 h de T.P. 10 h de Cours et 20 h de T.D.

TAB. 2.1 – Service statutaire effectué au département de Mesures Physiques de l'IUT.

2.3 Formation continue

J'interviens de manière épisodique dans différents établissements universitaires (IUT, Polytech' Clermont, UFR Sciences et Techniques) pour des actions de formation continue et d'encadrement de projet.

1991 : Enseignement intégré (C, T.D. et T.P.) sur les systèmes à microprocesseurs dans le cadre de la formation continue pour des techniciens de la banque de France (20 HTD).

1998 : Enseignement intégré sur les mesures automatisées pour la préparation à un concours interne de techniciens de l'entreprise M.S.D. Chibret de Clermont-Ferrand (30 HTD).

2000 : Automatique linéaire continue au C.U.S.T., cycle préparatoire du département Génie Electrique, formation à distance (8 HTD).

2002 : Automatisme des systèmes discrets au C.U.S.T., cycle préparatoire du département Génie Electrique, formation à distance (8 HTD).

2.4 Tâches d'intérêt collectif

Je participe activement aux tâches administratives des universités clermontoise.

1998-2001 : Membre élu représentant des Maîtres de Conférences au conseil du département de Mesures Physiques de l'I.U.T. d'Aubière, Université d'Auvergne.

1998-2000 : Membre extérieur nommé de la commission de spécialistes de l'Institut Français de Mécanique Avancée, sections 60 et 61 du C.N.U.

Depuis 2000 : Membre extérieur nommé de la commission de spécialistes de l'Université Blaise Pascal, section 61 du C.N.U.

Depuis 2001 : Membre élu de la commission de spécialistes de l'Université d'Auvergne, sections 28,61,63 du C.N.U.

2003-04 : Membre du comité de création de la licence professionnelle "Plasturgie et matériaux composite, option polymère dans l'environnement (ENVIPOL)" .

2007-08 : Membre du comité de création de la licence professionnelle "Sciences de la Mesure et du Contrôle (SIMCO)".

2008 : Membre élu du Conseil Scientifique de l'Université d'Auvergne.

3 Activités de recherche

3.1 Contexte

J'effectue mes travaux de recherche au LA.S.M.E.A., U.M.R. 6602 C.N.R.S. Université Blaise Pascal dans le GRoupe Automatique : VIsion et Robotique (GR.A.VI.R.) qui a pour objectif scientifique principal de développer des concepts théoriques et méthodologiques pour la perception et le contrôle de systèmes, et de valider ces concepts sur des démonstrateurs en situation réelle.

J'ai bénéficié d'une prime de recherche et d'encadrement doctoral attribuée lors de la campagne 2002 et d'une délégation de recherche au C.N.R.S. en 2004-05 et 2005-06.

Je bénéficie d'une prime de recherche et d'encadrement doctoral attribuée lors de la campagne 2006.

3.2 Encadrement scientifique

3.2.1 Post-doctorants

01/04/2006-30/09/2006 Baek Seung-Min, "Fusion de données pour la localisation d'un véhicule en environnement urbain en vue de son guidage automatique".

3.2.2 Doctorants

1998-2001 : Aufrère Romuald, "Reconnaissance et suivi de route par vision artificielle, application à l'aide à la conduite", thèse de l'université Blaise Pascal de Clermont-Ferrand soutenue le 20 juin 2001 (encadrement 35%).

2001-2006 : Laneurit Jean, "Fusion d'informations pour la localisation d'un véhicule", thèse de l'université Blaise Pascal de Clermont-Ferrand soutenue le 03 Juillet 2006 (encadrement 50%).

2004- 2007 : Tessier Cédric, "Système de localisation basé sur une stratégie de perception cognitive appliquée à la navigation autonome d'un robot mobile ", thèse de l'université Blaise Pascal de Clermont-Ferrand en cours de préparation (encadrement 33%), soutenue le 9 novembre 2007.

2002-2007 : Trujillo Noël, "Stratégie de perception pour la compréhension de scènes par une approche focalisante, application à la reconnaissance d'objets ", thèse de l'université Blaise Pascal de Clermont-Ferrand , soutenue le 13 décembre 2007 (encadrement 33%).

2004- : Karam Nadir, ”Localisation collective de robots communicants”, thèse de l’université Blaise Pascal de Clermont-Ferrand en cours de préparation (encadrement 33%).

2005- : Ngako-Pangop Laurence, ”Reconnaissance de piétons par perception multisensorielle ”, thèse de l’université Blaise Pascal de Clermont-Ferrand en cours de préparation (encadrement 33%).

2006- : Grimbichler David, ”Apports de la programmation parallèle pour la conception et l’optimisation de méthodes de perception multi sensorielle”, thèse de l’université Blaise Pascal de Clermont-Ferrand en cours de préparation (encadrement 50%).

3.2.3 Stagiaires D.E.A.

1998 : Aufrère Romuald, ”Suivi de route non marquée par vision monoculaire embarquée”, stage du D.E.A. Electronique et Systèmes de l’Université Blaise Pascal.

2000 : Chonier Frédéric, ”Localisation de véhicule sur route droite marquée par accumulation dans un espace de paramètres”, stage du D.E.A. Electronique et Systèmes de l’Université Blaise Pascal.

2001 : Laneurit Jean, ”Localisation d’un avion sur piste par vision monoculaire”, stage du D.E.A. Composants et Systèmes pour le Traitement de l’Information de l’Université Blaise Pascal.

2002 : Voisin Vincent, ”Localisation et suivi de route par coopération vision/G.P.S.”, stage du D.E.A. Composants et Systèmes pour le Traitement de l’Information de l’Université Blaise Pascal.

2004 : Karam Nadir, ”Suivi d’objet par Support Vector Machine. Application au suivi de véhicule”, stage du D.E.A. Composants et Systèmes pour le Traitement de l’Information de l’Université Blaise Pascal.

2005 : Gay-Bellile Vincent, ”Détection de piétons dans des images”, stage du master recherche Composants et Systèmes pour le Traitement de l’Information option VIsion pour la RObotique de l’Université Blaise Pascal.

3.3 Publications et communications

Je suis co-auteur de :
– 7 articles dans des revues à comité de lecture
– 30 communications à des conférences internationales avec actes et comité de sélection
– 3 communications à des conférences nationales avec actes et comité de sélection
– 3 communications à des colloques sans comité de sélection
Les références correspondantes sont maintenant détaillées.

3.3.1 Thèse

F. Chausse : *Reconstruction 3D de courbes paramétriques polynomiales par filtrage temporel. Approche par coopération vision par ordinateur / infographie. Application aux scènes routières.*

thèse de Doctorat de l'Université Blaise Pascal, Clermont-Ferrand, 21 Décembre 1994.

Directeur de thèse : Monsieur Christian THIBAUD, Professeur au C.U.S.T.

Rapporteurs : Monsieur André GAGALOWICZ, Directeur de recherche à l'I.N.R.I.A.

Monsieur Radu HORAUD, Directeur de recherche au C.N.R.S.

3.3.2 Articles dans des revues nationales et internationales à comité de lecture

R. Chapuis, A. Potelle, J.L. Brame, F. Chausse : *Real time vehicle trajectory supervision on highway*, International Journal of Robotics Research,14(6) :531-542, Dec 1995.

C. Rouquet, F. Chausse, R. Chapuis, P. Bonton : *Segmentation non supervisée d'images de scènes routières. Une approche multi-critère*, Traitement du Signal, 13(3), 1996.

R. Aufrère, R. Chapuis, F. Chausse : *A dynamic Vision algorithm to locate a vehicule on a non-structured Road*, International Journal of Robotics Research, 19(5), pp. 411-423, Mai 2000.

R. Aufrère, R. Chapuis, F. Chausse : *A model driven approach for real time road recognition*, International Journal of Machine Vision, 13, pp. 95-107, 2001.

R. Chapuis, R. Aufrère, F. Chausse : *Accurate road following and reconstruction by computer vision*, IEEE Transactions on Intelligent Transportation Systems, 3(4), pp. 261-270, Décembre 2002.

J. Laneurit, F. Chausse, and R. Chapuis : *Accurate vehicle positioning onto a numerical map*, International Journal of Control, Automation, and Systems, mars 2005, vol. 3, n˚ 1, pp.15-31.

F. Chausse, J. Laneurit, R. Chapuis : *Localisation d'un véhicule sur une carte routière précise*, Traitement du Signal, vol. 23, n˚2, avril 2006, pp.159-178.

3.3.3 Communications à des conférences internationales avec actes et comité de lecture

I. Coulibaly, P. Bonton, F. Chausse, C. Thibaud : *Estimating motion from image sequences of a multisensor*, Picture Coding Symposium, Lausane, March 1993.

P. Sayd, R. Chapuis, F. Chausse, C. Rouquet : *Non structured road following : an efficient monochromatic vision algorithm*, Advances in vehicle control and safety, Amiens, July 1-3 1998.

P. Sayd, R. Chapuis, R. Aufrére, F. Chausse : *A dynamic vision algorithm to recover the 3D shape of a non structured road*, I.E.E.E. International Conference on Intelligent Vehicles, Stuttgart, October 1998.

R. Chapuis, R. Aufrère, F. Chausse : *Recovering the 3D shape of a road by vision*, 7th International Conference on Image Processing and its Applications-IPA99, Manchester, UK, July 1999, pp. 686-690.

F. Chausse, R. Aufrère, R. Chapuis : *Implementation of a new vision based road following algorithm in the VELAC experimental car*, International Conference on Control Automation Robotics and Vision, December 2000, Singapore, CD file/paper/318.pdf.

R. Aufrère, R. Chapuis, F. Chausse : *Real time vision based road lane detection and tracking*, MVA 2000, November 2000, Tokyo Japan, pp.75-78.

R. Aufrère, R. Chapuis, F. Chausse : *A fast and robust vision based road following algorithm*, IV 2000, Dearborn, Michigan, USA, October 2000, pp.192-197.

F. Chausse, R. Aufrère, R. Chapuis : *Vision based vehicle trajectory supervision*, ITSC 2000, Dearborn, Michigan, USA, October 2000, pp.143-148.

F. Chausse, R. Aufrère, R. Chapuis : *Recovering the 3D Shape of a road by on-board monocular vision*, ICPR 2000, Barcelona, Spain, September 2000, vol. 1, pp. 325-328.

R. Aufrère, R. Chapuis, F. Chausse, J. Alizon : *Road sides recognition under unfriendly lightning conditions*, IV 2001, Tokyo, Japan, May 2001, pp. 13-18.

R. Chapuis, J. Laneurit, R. Aufrère, F. Chausse, T. Château : *Accurate vision based road tracker*, IV 2002, Versailles, France, Juin 2002, CD rec_papers/IV-103.pdf.

J. Laneurit, F. Chausse, R. Aufrère, R. Chapuis : *Vehicle control using image cues recognition*, International Sympopsium on Measurement and control in Robotics, Bourges, France, juin 2002.

F. Chausse, V. Voisin, J. Laneurit, R. Chapuis : *Centimetric localisation of a vehicleby combining vision and low cost GPS*, MVA 2002, Nara, Japon, décembre 2002, pp. 538-541.

M. Bertozzi, A. Broggi, R. Chapuis, F. Chausse, A. Fascioli, A. Tibaldi : *Shape-based pedestrian detection and localization*, actes de la conférence internationale Intelligent Transportation Systems 2003, Shangai, Chine, Octobre 2003, pp. 328–333.

M. Bertozzi, A. Broggi, A. Fascioli, A. Tibaldi, R. Chapuis, F. Chausse : *Pedestrian Localization and Tracking System with Kalman Filtering*, Intelligent Vehicle Symposium, Parma, Italie, Juin 2004, pp 584-589.

F. Chausse, N.Trujillo, R. Chapuis and M. Naranjo : *Object recognition by model based focused vision*, International Conference on Advances in Intelligent Systems - Theory and Applications, Luxembourg, November 15-18, 2004 (Actes sur CD-ROM, fichier 172.pdf, 6 Pages).

N.Trujillo, R. Chapuis, F. Chausse, C. Blanc : *On road simultaneous vehicle recognition and localization by model based focused vision*, IAPR Conference on Machine Vision and Applications, Tsukuba, Japon, 16-18 mai 2005, pp 388-391

F. Chausse, J. Laneurit, R. Chapuis : *Vehicle localization on a digital map using particles filtering, Intelligent Vehicle Symposium*, Las Vegas, USA, 6-8 juin 2005.

C. Tessier, C. Cariou, C. Debain, R. Chapuis, F. Chausse, C. Rousset : *A Real-Time, Multi-Sensor Architecture for fusion of delayed observations : Application to Vehicle Localisation*, 9th International IEEE Conference on Intelligent Transportation Systems, ITSC 2006, Toronto, Canada, 18-20 Sept. 2006.

C. Tessier, C. Debain, R. Chapuis, F. Chausse : *Fusion of active detections for outdoor vehicle guidance*, The 9th International Conference on Information Fusion, Florence (Italy), 10-13 July 2006, IEEE Catalog No. :06EX1311C, ISBN 0-9721844-6-5.

C. Tessier, C. Debain, R. Chapuis, F. Chausse : *Active Perception Strategy for Vehicle Localisation and Guidance*, the 2nd IEEE International Conferences on Cybernetics & Intelligent

Systems and Robotics, Automation & Mechatronics (CIS-RAM 2006), 7-9 June, 2006, Bangkok, Thailand, IEEE Catalog No. :06EX1247C, ISBN 1-4244-0025-2.

T. Chateau, V. Gay-Belille, F. Chausse, J.-T. Lapresté : *Real-Time Tracking with Classifiers*, 14 pages, WDV 2006 - WDV Workshop on Dynamical Vision at ECCV2006, Graz, Austria May 13, 2006.

N. Karam, F. Chausse, R. Aufrère, R. Chapuis : *Cooperative Multi-Vehicle Localization*, in proceedings, IEEE IV 2006, Tokyo Japan, June 2006. pp.564-570.

N. Karam, F. Chausse, R. Aufrère, R. Chapuis : *Collective Localization of Communicant Vehicles Applied to Collision Avoidance*, in proceedings, IEEE ITSC 2006, Toronto Canada, September 2006.

N. Karam, F. Chausse, R. Aufrère, R. Chapuis : *Localization of a Group of Communicating Vehicles by State Exchange*, in proceedings, IEEE IROS 2006, Beijing China, October 2006.

L. Ngako Pangop, F. Chausse, S. Cornou, R. Chapuis : *Feature-Based Multisensor Fusion Using Bayes Formula for Pedestrian Classification in Outdoor Environments*, in proceedings, IEEE IV 2007, Istanbul Turquie, 13-15 Juin 2007.

C. Tessier, M. Berducat, R. Chapuis, F. Chausse : *A New Landmark and Sensor Selection Method for Vehicle Localization and Guidance*, in proceedings, IEEE IV 2007, Istanbul Turquie, 13-15 Juin 2007.

L. Ngako Pangop, F. Chausse, S. Cornou, R. Chapuis : *Adaptative batesian combination of features from laser scanner and camera for pedestrian detection*, in proceedings, IAV'07 - 6ème IFAC Symposium on Intelligent Autonomous Vehicles, Toulouse, France, 3-5 September 2007

C. Tessier, C. Debain, R. Chapuis, F. Chausse : *Simultaneous Landmarks Detection and Data Association in Noisy Environment for Map Aided Localization*, in proceedings, IEEE International Conference on Intelligent Robots and Systems IROS, San Diego, CA, USA, Oct 29 - 2 Nov, 2007.

N. Trujillo, R. Chapuis, F. Chausse, M. Naranjo M. : *Object Recognition : A Focused Vision Based Approach*, in proceedings, International Symposium on Visual Computing, Lake Tahoe, Nevada/California, USA, Nov. 26-28, 2007

3.3.4 Communications écrites à des conférences nationales avec actes et comité de lecture

F. Chausse, C. Thibaud, R. Chapuis, A. Potelle : *Reconstruction récursive de courbes gauches à partir de leurs projections perspectives*, neuviéme congrés AFCET RFIA, Paris, 11-14 Janvier 1994.

R. Aufrère, R. Chapuis, F. Chausse : *Détection et suivi de voie de circulation par vision temps réel*, 18éme colloque ANRT Imagerie Rapide et Photonique, Paris, France, mai 2000, pp. 166-171.

R. Aufrère, R. Chapuis, F. Chausse : *Détection précise de bords de route par vison monoculaire embarquée*, RFIA 2000, Paris, France, février 2000, vol. 2, pp. 229-238.

3.3.5 Colloques sans comité de sélection

R. Aufrère, R. Chapuis, F. Chausse : *Détection et suivi de bords de routes en temps réel par vison artificielle*, ORASIS, Cahors, France, juin 2001, pp. 183-192.

Laneurit J., Chausse F., Aufrère R., Chapuis R. : *Localisation absolue d'un véhicule dans un environnement cartographié* , pp. 217-225, Innovation Technologique pour les Transport Terrestres, Colloque TILT, Lille France, 2-4 décembre 2003.

N. Trujillo, F. Chausse R. Chapuis : *Reconnaissance d'objets par vision focalisée*, Colloque ORASIS'05, Fournols, France, 24-27 mai 2005 (Actes sur CD ROM, fichier 016.pdf, 6 pages, non numérotées).

3.4 Collaborations de recherche

Les résultats de mes travaux de recherche ont été valorisés dans le cadre de collaborations industrielles ou avec des universités étrangères auxquelles j'ai activement contribué.

1997-98 : Contrat avec la société Aérospatiale (200 kF) : détection de voies non structurées (chemins, etc. . .) dans des séquences d'images pour l'aide au déplacement de robots militaires.

1999-2000 : Contrat avec la société Aérospatiale Matra Missile (230kF) : aide au guidage d'avions sur piste par analyse d'image.

2000-01 : Contrat avec la société Peugeot S.A. (50kF) : faisabilité de la mesure d'adhérence des revêtements routiers par analyse d'images.

2001 : Contrat avec la D.G.A. (15k€) : programmes de traitements d'images pour la navigation de véhicule robotisés ;

lot 1 : détection de bord de routes et de chemins peu ou non structurés sur images monochromes, avec une approche de type région fondée sur une classification à partir d'attributs monochromes.

lot 4 : détection de bord de routes et de chemins peu ou non structurés sur images monochromes, avec une approche de type contour fondée sur la recherche récursive de contours en lignes brisées dans des zones d'intérêt dynamiques.

2001-03 : Collaboration avec le C.E.M.A.G.R.E.F. dans le cadre de l'axe prioritaire Machines et Systèmes Performants et Intelligents du Contrat de Plan Etat Région Auvergne : perception et commande pour l'aide à la conduite des véhicules routiers et agricoles.

2001-03 : Collaboration dans le cadre d'un Programme d'Action Intégré Européen avec le Groupe Vision Artificielle du Dipartimento di Ingegneria dell'Informazione de l'université de Parme (Italie) dirigé par le Professeur Alberto Broggi : détection et suivi de piétons dans des séquences d'images de scènes routières.

2002-03 : Contrat avec la société THALES Systèmes Aéroportés (30k€) : Détection des bords de voie de route ou de chemin par analyse d'image pour l'aide à la navigation de véhicules.

2003-04 : Collaboration avec le C.E.M.A.G.R.E.F. (30k€) dans le cadre d'un contrat avec la

société ECA pour le programme d'étude en amont MiniROC de la DGA : localisation précise d'un robot dans un environnement cartographié par combinaison d'informations proprioceptives, G.P.S. et vision dans le but d'assurer une navigation autonome.

2004-07 : Participation au projet MobiVip du PREDIT (20k€) : fusion de données pour la localisation d'un véhicule en environnement urbain.

2006-09 : Responsable au LASMEA du projet ANR LOVe (310k€) : Logiciel d'Observation des Vulnérables - Détection reconnaissance et suivi de piétons.

3.5 Groupes de recherche nationaux

Je participe depuis 2004 aux activités du groupe de recherche "Fusion de données pour la localisation" (FUDOLO). L'objectif est de réaliser la comparaison de méthodes de fusion d'information (méthodes bayésiennes, méthodes ensemblistes, ...) pour la localisation d'un véhicule. Cette comparaison est quantitative et s'effectue sur différents ensembles de données multicapteurs acquises en situation réelle.

Deuxième partie

Travaux de recherche

Introduction

Les travaux de recherche présentés dans cette partie ont été menés sur la période 1995-2006 au Laboratoire des Sciences et Matériaux pour l'Electronique, et d'Automatique (LASMEA) dans le Groupe Automatique, Vision et Robotique (GRAVIR). Les objectifs de ce groupe sont de développer des concepts théoriques et méthodologiques pour la perception et le contrôle de systèmes, et de valider ces derniers sur des démonstrateurs en situation réelle.

Ces activités ont trouvé depuis une vingtaine d'années un champ d'application particulier dans le domaine du développement de systèmes d'assistance à la conduite automobile. Ce domaine a par la suite été étendu à des applications de robotique mobile en environnement extérieur.

Les robots mobiles évoluent constamment dans un environnement qui n'est pas statique : ils se déplacent dans une scène dynamique. Pour interagir avec cette scène, ils doivent la "comprendre" c'est-à-dire connaître à chaque instant son état en recueillant des informations grâce à des capteurs. C'est pourquoi le groupe GRAVIR travaille au développement de systèmes de perception. C'est à ce thème de recherche en particulier que sont rattachés les travaux exposés dans ce document.

Les modèles de systèmes de perception pour la robotique ont été historiquement développés dans le cadre de la vision par ordinateur. La quantité importante d'informations contenue dans les images a en effet rendu nécessaire l'écriture d'algorithmes mettant en oeuvre une stratégie d'analyse adaptée pour extraire efficacement l'information utile à la compréhension de la scène. Le modèle séquentiel de Marr [54] a donné naissance à la vision par ordinateur qui ne s'est plus réduit au simple traitement des images. Lui ont succédé le modèle de la vision active présenté par Aloimonos [88] puis celui de perception active introduit par Bacsy [5] et enfin celui de la vision attentionelle que Tsotsos a formalisé [80]. L'idée est de mettre en oeuvre un processus de contrôle qui guide l'analyse des images vers l'information intéressante. L'analyse exhaustive de l'image n'est pas nécessaire. Le contrôle du processus attentionnel peut être guidé par les caractéristiques (*bottom-up*), par le but (*top-down*) ou par les deux.

C'est dans ce cadre formel que s'inscrivent les travaux présentés dans ce mémoire. Un système de perception attentionnelle est proposé et mis en œuvre dans un contexte multisensoriel. Un robot est équipé de différents types de capteurs extéroceptifs (ex. : caméra vidéo, télémètre LASER) qui recueillent des données sur la scène environnante. La compréhension consiste à instancier un modèle de la scène à partir des informations issues des capteurs. Ce modèle est une représentation synthétique de la scène, une carte dynamique de l'environnement du robot qui regroupe les informations utiles à la tâche robotique. Ces informations sont principalement de nature géométrique (pose du robot et des obstacles) afin d'assurer la tâche de localisation puis de navigation, dans l'objectif important de contrôler la mobilité du robot. Elles peuvent être complétées par d'autres types d'informations non directement utiles à une tâche particulière que le robot doit accomplir (ex. : apparence visuelle des objets) mais qui permettent de faciliter

la compréhension.

Le système de perception attentionnelle proposé est dans son principe de type *top-down* : la compréhension d'une scène dynamique est guidée par sa représentation synthétique. La reconnaissance individuelle des objets qui la composent permet d'instancier la scène à l'instant présent. En retour, la reconnaissance pour l'instant suivant est guidée par le modèle de la scène prédite grâce à une procédure d'évolution dynamique. Ce processus focalise la recherche des caractéristiques de la scène à l'intérieur de zones d'intérêt qui correspondent à des sous-espaces de l'espace des données de chaque type de capteur. L'observation d'une caractéristique compatible avec l'état courant de la scène entraîne la remise à jour de la représentation synthétique qui s'affine progressivement. La représentation synthétique de la scène inclut la prise en compte du lien statistique entre ses composantes. Ainsi, la reconnaissance d'une partie de la scène dans les données capteur focalise un peu plus celle des autres parties, même si ces dernières n'ont pas été directement observées.

Ce principe de **focalisation progressive** de la recherche dans l'espace des observations est le cœur du modèle de perception attentionnelle proposé. Il a été introduit entre 1998 et 2000 par Roland CHAPUIS qui en a présenté le principe et la première mise en œuvre dans son mémoire d'Habilitation à Diriger des Recherches [71]. Il fait l'originalité de mes travaux de recherches et en constitue la ligne directrice. Il a été mis en oeuvre dans deux applications particulières : la localisation d'un robot dans son environnement et la détection/reconnaissance d'objet mobiles. De plus que mon travail se caractérise par la volonté systématique de mise en œuvre expérimentale des méthodes de perception sur des robots démonstrateurs.

Cette deuxième partie du mémoire qui décrit mes activités de recherche est constituée de trois chapitres. Le premier analyse quelques publications des trente dernières années qui jalonnent les avancées marquantes sur les systèmes de vision. Il est destiné à définir ce qu'est un système de perception et à positionner celui qui est proposé ici dans un contexte moins historique (même si les travaux référencés sont décrits dans l'ordre chronologique) que méthodologique. Le deuxième et le troisième chapitre abordent respectivement les deux types d'applications pour lesquelles ce système de perception est mis en œuvre : la localisation d'un robot et la perception des obstacles. Ils détaillent les contributions apportées dans cinq thèses que j'ai co-encadrées.

4 Systèmes de perception

4.1 Introduction

Un système est communément défini [1] comme un ensemble d'éléments matériels ou non en relation les uns avec les autres et formant un tout. Dans le contexte de la perception pour la robotique, un système est constitué d'éléments matériels et logiciels qui concourent à la réalisation d'un but, d'une tâche robotique. Parmi les éléments matériels, les capteurs extéroceptifs recueillent des informations sur la scène environnant le robot.

Bien qu'il faille se garder de négliger les aspects capteurs et architectures de calcul, ils ne sont pas l'objet du travail présenté ici car ils constituent des axes de recherche spécifiques. Or, même si pour soutenir l'intérêt d'un système de perception qui se veut global, l'affichage d'une scission trop nette entre les différentes composantes du problème peut paraître contradictoire, l'efficacité de la recherche nécessite un minimum de spécialisation. Mon travail est plutôt orienté vers les aspects algorithmiques de la perception qui sont donc les seuls abordés dans ce document.

Les informations se présentent en sortie des capteurs sous la forme de données ”brutes” impossibles à utiliser telles quelles pour réaliser la tâche robotique. Elles doivent être traitées pour en extraire une représentation synthétique de la scène utile à l'accomplissement de la tâche. Définir l'enchaînement des traitements nécessaires est d'une très grande difficulté car l'information utile est mêlée à une importante quantité de données sans intérêt.

Ainsi, il s'agit de mettre en œuvre une stratégie de traitement visant à construire une représentation symbolique efficace de la scène à partir des données directement issues des capteurs. Le terme efficace n'a bien entendu de sens qu'au regard du but à atteindre. Un système de perception doit donc établir une relation entre les données capteurs, parfois appelées stimuli, et l'information synthétique qui modélise l'état à un moment donné de la connaissance de la scène. Il est caractérisé par la stratégie mise en œuvre pour établir ce lien.

Les stratégies de perception ont été définies au fil du temps dans le domaine de l'analyse d'image. Dans ce cas, les systèmes de perception sont constitués d'un seul capteur. Ceci ne nuit en rien à leur généricité. Leur étude, au travers d'un nombre finalement assez faible de publications clés, donne bien sûr un aperçu de l'évolution des systèmes perception artificielle mais permet surtout de poser parfaitement le problème scientifique et aide ainsi à fixer les idées sur le positionnement du travail présenté dans ce mémoire.

4.1.1 Le modèle de Marr : la vision par ordinateur

Une image est un signal bidimensionnel, et à ce titre peut être traitée par les opérateurs classiques de traitement du signal (filtrage, amplification sélective d'une composante, etc.) pour en

extraire l'information utile. L'interprétation de cette information dans un but donné ne présente pas de problème. . .C'est du moins la remarque qui a cours pendant la deuxième moitié des années 60 à l'époque des débuts de l'analyse artificielle des images. L'expérience a ensuite montré et montre encore après 40 ans d'une recherche internationale particulièrement active que ce problème est très compliqué.

Mathématicien et spécialiste de neurosciences, D. Marr l'a compris dès la fin des années 80. Il s'est efforcé de construire une théorie, celle de la vison par ordinateur [54], qui est à la base du tout premier système de perception. Le schéma bien connu (figure 4.1) montre un traitement de l'information à trois niveaux d'abstraction (bas, moyen et haut niveau). Le bas niveau met en œuvre des opérateurs de traitement du signal pour extraire de l'image des **caractéristiques** (ex. : les contours) constituant l'information utile. Au moyen niveau ces caractéristiques sont regroupée entre elle selon des critères d'interprétation donné (ex. : tel groupe de points de contour peut constituer une droite). Un modèle de la scène est construit au haut niveau à partir de ces groupes de caractéristiques.

Ce schéma fait apparaître un lien direct entre une image et une représentation, un modèle de la scène observée, modèle structuré sous forme de surfaces élémentaires. Un niveau de rang inférieur transmet l'information en entrée du niveau supérieur et il n'y a pas d'autre lien entre les niveaux que cette organisation hiérarchique. C'est un système en boucle-ouverte (pour prendre le langage des automaticiens) donc où la sortie n'exerce aucun contrôle sur l'entrée ce qui laisse au module bas niveau la lourde responsabilité de ne pas se tromper dans l'extraction des caractéristiques utiles.

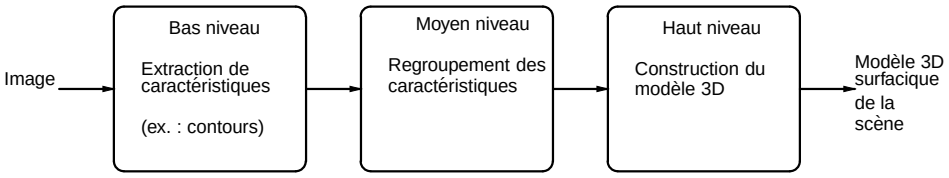


FIG. 4.1 – Schéma du système de perception proposé par Marr

4.1.2 Le modèle d'Aloimonos : la vision active

Y. Aloimonos [88] met en avant les limites du modèle de Marr en précisant qu'il pose mal le problème puisque Marr prétend construire un modèle 3D de la scène à partir d'une image 2D. Sans remettre en cause son système de perception, Aloimonos souligne que Marr oublie un élément important dans son système de perception : l'observateur et en particulier le fait qu'il se déplace.

Avec la vision active, Aloimonos montre que la prise en compte explicite du mouvement de la caméra permet de résoudre l'ambiguïté 3D/2D. Ce modèle reste en boucle ouverte. En effet, même si le déplacement de l'observateur est pris en compte, il n'est pas directement contrôlé par le système de perception mais subit les évolutions requises par l'application. L'idée est donc de tirer partie du mouvement sans toutefois le contrôler.

En outre cette approche suppose qu'il est possible de connaître le déplacement de la caméra, ce qui n'est pas un problème simple et laisse en suspens une question importante : que faire lorsque l'application ne nécessite pas de déplacement de la caméra ?

4.1.3 Le modèle de Bajcsy : la perception active

R. Bajcsy [5] va un peu plus loin qu'Aloimonos en précisant que dans l'objectif d'interpréter une scène, la perception est une activité exploratoire qui nécessite de définir le déplacement de l'observateur non pas à partir des seuls objectifs de l'application mais également à partir des objectifs de la tâche de perception.

C'est le premier système de perception en boucle fermée où une stratégie est mise en œuvre pour contrôler la perception et que l'auteur définit très bien elle-même :

« *Par conséquent, le problème de la perception active peut être posé comme un problème de stratégie de contrôle appliquée au processus d'acquisition des données et qui dépend de l'état courant de l'interprétation des données et du but ou de la tâche de l'application. A la question est-ce que la perception active n'est qu'une application de la théorie du contrôle ? Notre réponse est non, tout au moins pas dans sa version simple.*
Voici pourquoi :

1. *la boucle de retour ne s'applique pas aux seules données capteurs mais aussi à des informations sensorielles plus complexes, i.e. des caractéristiques de nature variée extraites des images, incluant des relations entre caractéristiques.*
2. *La boucle de retour est dépendante de modèles de connaissance a priori qui sont des mélanges d'informations numériques ou paramétriques et symboliques.*

»

Aux **approches modulaires** (avec ou sans prise en compte du mouvement) proposées par Marr et Aloimonos, Bajcsy oppose clairement une **approche système** qui nécessite de définir une stratégie globale de perception.

Bajcsy fait également la remarque importante que l'effort de recherche ne devrait pas se concentrer sur l'optimisation des opérateurs d'extraction des caractéristiques qui seront toujours imparfaits mais sur la prise en compte de cette imperfection dans le système de perception.

Ainsi, selon cette approche, l'idée de la perception active laisse entrevoir, au travers de la nécessité d'un processus de contrôle de la perception, celle d'orienter la prise d'information vers les parties importantes de la scène au sens du but à atteindre. Ceci laisse entrevoir la nécessité d'un processus de focalisation de l'attention en perception.

L'approche perception active de Bajcsy a été utilisée par Ballard et Brown dans leur modèle de la vision animée [6] développé entre 1989 et 1992 et qui en constitue une déclinaison originale. Ce modèle est fondée sur le contrôle du point de vue (*gaze control*) inspiré des mouvements automatiques de l'œil humain (saccade oculaire). La vision animée n'apporte pas d'innovation sur le fond au principe de la vision active.

4.2 Perception attentionnelle

R. Bajscy accompagne sa théorie de remarques liées au système visuel humain ("We do not just see, we look") que la communauté de la perception artificielle (en particulier visuelle) n'a jamais prétendu égaler ni même imiter mais dont elle s'inspire depuis longtemps. Parmi les processus qui forment le système de perception naturel, la focalisation de l'attention témoigne du

rôle actif de l'observateur dans l'acte perceptif. Le but est de réduire au plus tôt la quantité d'information brute à traiter pour transmettre le plus rapidement possible les données importantes à un niveau décisionnel.

La perception attentionnelle nécessite un processus de contrôle qui met en œuvre la stratégie de gestion des informations (choix des modèles, des opérateurs, des données . . .). Trois types de mécanismes de contrôle existent :

1. Contrôle ascendant (*bottom-up*) : indépendant du but à atteindre le contrôle ascendant guide le processus attentionnel à partir des caractéristiques les plus saillantes extraites des données capteurs. Les données ainsi extraites sont ensuite transmises à un processus de plus haut niveau (reconnaissance, analyse, etc. . .) qui détermine les informations nécessaires à l'application.

2. Contrôle descendant (*top-down*) : l'attention est guidée par le but à atteindre qui est modélisé par la représentation synthétique de la scène perçue. Le contrôle descendant se caractérise notamment par la définition de zones d'intérêt des données capteurs dans lesquelles les caractéristiques importantes de la scène perçue sont susceptibles de se trouver. Il est par exemple possible de prendre en compte des informations contextuelles pour contraindre la détection des caractéristiques.

3. Contrôle hybride : les deux processus précédents interagissent dans le système de perception.

J.K. Tsotsos [80] fait de l'attention la principale caractéristique d'un système de perception (pour lui limité à la vision, mais le principe peut se généraliser à tout type de capteur) avec pour objectif la diminution de la complexité combinatoire d'une approche qui analyse exhaustivement les données capteurs. Il définit la focalisation de l'attention à plusieurs niveaux :
– focalisation dans l'espace géométrique du capteur : contrôle du champ de "vue" du capteur donc de sa pose comme dans la perception active,
– focalisation dans l'espace des caractéristiques : contrôle de l'analyse des données capteurs en termes de zone d'intérêt, de réglage (résolution, seuils, etc. . .) des opérateurs de calculs des caractéristiques,
– focalisation dans l'espace des modèles : choix du modèle de la scène,
– focalisation dans l'espace des décisions : choix d'événements, d'objets ou de tâches à accomplir pour atteindre l'objectif de l'application.

Focaliser l'attention de la perception relève donc de plusieurs choix permettant de sélectionner,
– la caractéristique à observer ("quoi"),
– la zone d'intérêt dans l'espace des capteurs dans laquelle observer cette caractéristique ("où"),
– l'opérateur le plus à même de réaliser cette observation ("comment").

Le choix s'opère en fonction de critères basés sur la pertinence, la facilité, l'aptitude des opérateurs, la vraisemblance ou la complétude. Au final, le système de perception enchaîne une suite d'opérations à même de fournir l'information utile à l'application en optimisant à chaque étape un compromis utilité/coût. Un processus attentionnel s'accompagne d'un mécanisme d'adaptation, destinée à maintenir sa cohérence, qui permet la remise en cause d'une étape de l'analyse perceptive .

4.2.1 Bilan

Cette rapide présentation des systèmes de perception et en particulier les systèmes attentionnels est destinée à exposer le cadre formel du travail présenté dans ce mémoire. Une synthèse plus complète est présentée par J.-M Jolion et ses co-auteurs ([46], en particulier les chapitres 4 et 7). Les systèmes de perception attentionnels de type *bottom-up* tel que le concept de carte de saillance introduit par Koch et Ulmann au milieu des années 80 [17] et qui a été 15 à 20 ans plus tard largement diffusés notamment par Itti [42] ou Christensen [69], ne sont qu'évoqués et pas détaillés car l'approche proposée et décrite n'est pas de ce type.

En effet, ce mémoire décrit un système attentionnel de type *top-down* et les points de repères bibliographiques principalement cités se limitent à cette classe d'approches. De même certaines déclinaisons de la perception active menées par Draper [27] ou par Davison [24, 44] ne servent pas de référence dans cette partie car elles n'apportent pas de modifications significatives du concept. Que cela ne soit pas interprété comme une minimisation de la portée de ces travaux qui restent remarquables par la mise en œuvre originale de la perception active.

Les travaux pionniers de Marr, d'Aloimonos et de Bacsy jalonnent l'évolution des réflexions de la communauté sur la perception artificielle. Ruzena Bacsy a fait il y a 20 ans des remarques très visionnaires pour l'époque qu'elle a réitérées à San Diego (USA) fin 2007 à l'occasion d'une conférence invitée au colloque IROS (Intelligent Robots and Systems). Ces remarques introduisent la notion de focalisation de l'attention en perception pour la robotique. Tsotsos à formalisé et développé ensuite cet aspect.

Le système de perception proposé dans ce mémoire est fortement appuyé sur les concepts utilisés dans ces travaux. Il s'agit d'un système attentionnel basé sur un processus de focalisation dans l'espace des caractéristiques tel que le définit Tsotsos. Il met en œuvre les remarques de Bacsy puisqu'il repose sur un processus de contrôle s'appuyant sur l'état courant de la perception pour orienter l'analyse des données issues des capteurs afin d'en extraire les informations utiles à la tâche robotique.

Ce système est présenté dans la section suivante.

4.3 Modèle de perception attentionnelle proposé

L'objectif de cette section est d'exposer les principes du système de perception attentionnel que nous proposons et autour duquel s'organisent les applications de perception pour la robotique mobile présentées ensuite. Ce système est principalement de type *top-down* bien qu'il intègre aussi pour certaines applications un processus *bottom-up* : les hypothèses de perception sont dirigées par la représentation synthétique de la scène. Cette représentation est un modèle statistique qui prend en compte les relations entre les caractéristiques des différentes parties de la scène.

4.3.1 Représentation synthétique de la scène

Structure globale du modèle

Le modèle qui guide la perception est constitué d'un vecteur d'état $\underline{X}$ dont les N éléments x_i $(i = 1..N)$ sont des paramètres utiles à l'application. Ils sont éventuellement complétés de paramètres d'importance secondaire pour la tâche robotique mais possédant un fort caractère discriminant ce qui les rend à même de faciliter la perception. La scène est a priori dynamique donc $\underline{X}$ évolue dans le temps. L'objectif de la perception est de mettre à jour ce vecteur d'état à partir de l'analyse des données en provenance des capteurs.

Au vecteur d'état est associé sa matrice de covariance $\mathbf{C_X}$ qui fait intégralement partie du modèle. Il s'agit donc d'une représentation statistique de la scène grâce à laquelle le système de perception est à même de prendre en compte la dispersion possible des paramètres (éléments diagonaux de la matrice de covariance) et les liens statistiques qui les unissent (covariances). A un instant donné, la scène est donc modélisée par la réalisation d'un vecteur aléatoire $\underline{X}$ qui suit la loi normale $\mathcal{N}(\underline{X}, \mathbf{C_X})$. La scène est donc a priori une instance unimodale de ce modèle.

Les objets

Pour être tout à fait général, il est nécessaire de considérer que les éléments x_i peuvent être eux mêmes des vecteurs. Ils sont alors notés $\underline{x_i}$ et sont associés à leur matrice de covariance $\mathbf{C_{xi}}$. Chaque vecteur $\underline{x_i}$ définit donc un sous-espace de l'espace d'état. Ainsi se dessine une représentation de la scène (figure 4.2) où chaque sous-espace de l'espace d'état définit un objet. A partir de chaque sous-espace peuvent être définis de la même manière d'autres sous-espaces modélisant d'autres objets. Une structure hiérarchique se dessine alors dans laquelle à chaque niveau un objet est relié à son objet "père" et à tous ses objets "enfants". Par exemple sur la figure 4.2, l'objet O_{11} a pour père l'objet racine O_1 et possède deux objets "enfants" O_{111} et O_{112}.

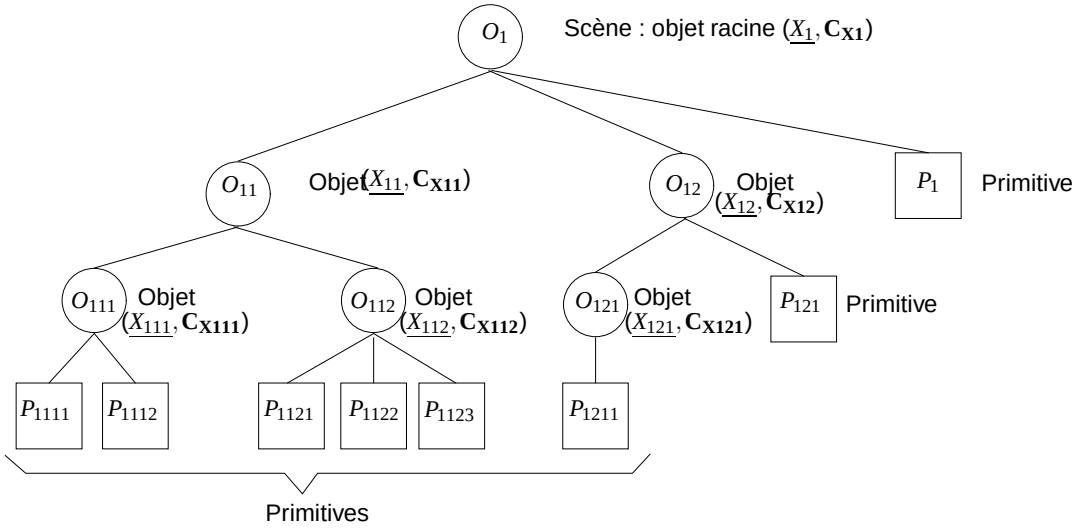


FIG. 4.2 – Exemple du modèle hiérarchique d'une scène

A chaque objet sont associées une fonction de mise à jour de son père M_i et une fonction de focalisation sur ses enfants F_i. Cette dernière est complétée d'une fonction de coût χ_i qui

évalue l'impact a priori sur le processus de perception de la focalisation sur tel ou tel enfant et une fonction d'utilité G_i qui détermine a priori également le bénéfice que la perception peut retirer en se focalisant sur l'un ou l'autre des enfants.

Les primitives

Le découpage hiérarchique se termine au niveau du capteur : la couche la plus basse de l'arbre est constituée d'enfants particuliers qui sont des primitives P_i extraites des données brutes par la fonction de focalisation qui leur est associée et qui prend alors le nom de détecteur.

Comme tous les objets, une primitive est modélisée par un vecteur d'état $\underline{X_{Pi}}$ et par sa matrice de covariance $\mathbf{C_{XPi}}$. Les éléments de $\underline{X_{Pi}}$ sont les **caractéristiques** de la primitive. Le détecteur est un algorithme de traitement des données capteur qui détermine les caractéristiques.

A titre d'exemple, dans le cas où le capteur est une caméra embarquée dans le robot, une primitive peut être constituée de caractéristiques telles que des paramètres de texture, des contours, les moments statistiques de la luminance, des points d'intérêt (points de Harris) etc. . .

Une primitive est donc un vecteur multi-caractéristiques décrivant localement la représentation d'un objet dans les données capteur.

4.3.2 Apprentissage

Il est possible de considérer *a priori* un état moyen de la scène associé à un intervalle de dispersion représenté par les covariances. Cet état initial représente les configurations réalistes admissibles pour le modèle de la scène. Sa prise en compte constitue le premier niveau de focalisation de la perception qui est appliqué a priori.

La détermination de cet état initial constitue une étape d'apprentissage dans le processus de perception de la scène. Elle consiste à donner une valeur initiale réaliste à chacun des éléments de chaque sous-espace $\underline{x_i}$ de l'espace d'état $\underline{X}$ et à la matrice de covariance $\mathbf{C_{xi}}$ qui lui est associée. Ceci définit le niveau $p = 0$ de la reconnaissance et l'état prend alors la valeur $(\underline{x_i}(0), \mathbf{C_{Xi}}(0))$.

Il n'y a pas de méthode générique pour réaliser l'apprentissage. Il peut être calculé à partir d'un modèle de connaissance de la scène (voir l'application de localisation d'un véhicule sur sa voie de circulation, section 5.1), ou bien être constitué d'une base de données (exemple : utilisation d'un SIG[1] dans l'application de localisation et guidage d'un robot mobile autonome, section 5.4), ou encore être obtenu expérimentalement par l'analyse d'une base de données d'apprentissage (ex. : reconnaissance des objets par vision focalisée, section 6.2). Une autre possibilité consiste à utiliser des contraintes (notamment géométriques) imposées par l'application (ex. : perception multisensorielle des piétons, section 6.5).

4.3.3 Evolution du processus attentionnel : la focalisation progressive

Pour un objet à un niveau hiérarchique donné, le système de perception oriente l'analyse sur celui de ses enfants qui présente le meilleur compromis coût/utilité. Ce processus se pour-

[1] Système d'Information Géographique

suit récursivement jusqu'au dernier niveau où la primitive de meilleur rapport utilité/coût est sélectionnée. Une suite d'**hypothèses de perception** s'enchaîne donc de l'objet racine (la scène) vers une primitive.

A chaque itération, l'état courant d'un objet (modélisé par son vecteur d'état $\underline{x_i}$ et sa matrice de covariance $\mathbf{C_{xi}}$) est utilisé par la fonction de focalisation pour délimiter dans l'espace d'état de ses enfants un sous-espace compatible avec l'hypothèse courante.

Arrivé au niveau d'une primitive, ce sous-espace correspond à une zone d'intérêt dans l'espace des caractéristiques de la primitive. A titre d'exemple et pour fixer les idées, dans le cas d'une primitive extraite d'une image vidéo, cette zone d'intérêt est une imagette dans laquelle une détection bas niveau (contour, texture, transformée en ondelette) s'applique. Cet exemple souligne que, pour une hypothèse donnée, le système de perception permet de ne pas effectuer la détection bas-niveau sur la totalité de l'image, ce qui accélère le traitement et améliore le rapport signal sur bruit. Cette remarque peut se généraliser à l'ensemble du processus : ainsi pour chaque type de primitive correspondant à un capteur donné, une zone d'intérêt peut être définie dans laquelle un traitement bas-niveau est ensuite appliqué pour calculer les caractéristiques de la primitive.

L'ensemble du réseau des hypothèses possibles s'organise sous la forme d'un **arbre d'interprétation**. Une scène est a priori très complexe et le parcours exhaustif de l'arbre à la recherche de la meilleure chaîne d'hypothèses de perception n'est pas envisageable car de complexité exponentielle. La méthode proposée apporte une solution à ce problème car elle permet d'élaguer au plus tôt les branches inutiles. Ce principe n'est pas nouveau (utilisé par Grimson dès 1987[40]) ; en revanche son implémentation dans le cadre de l'analyse de scène est originale.

L'algorithme de perception pour un objet O donné suit les étapes suivantes :

1. Pour chaque enfant O_e de O non encore traité :
   – Application de la fonction de focalisation de O_e
   – Calcul de l'utilité et du coût de O_e
2. sélection de Oe_{min} : l'enfant de O de meilleur compromis utilité/coût non encore pris en compte
3. Si la fonction de détection de Oe_{min} retourne un résultat compatible, l'état de O est mis à jour à partir de ce résultat
4. Sinon retour à l'étape 2

La description ci-dessus ne fait pas apparaître le caractère récursif de l'algorithme qui peut être appréhendé en considérant que ce même processus en 4 étapes est appliqué à l'enfant Oe_{min} puis à son "meilleur" enfant jusqu'à une primitive.

La mise à jour est typiquement effectuée par un filtre de Kalman étendu (EKF). La détection d'une primitive dans les données capteur déclenche la mise à jour de son objet père puis celle du père de son père et ce jusqu'à l'objet racine. Concrètement c'est le vecteur d'état et sa matrice de covariance associés à chaque objet qui sont mis à jour par l'EKF à chaque itération. Ceci entraîne la diminution des variances des vecteurs d'état de chaque objet du modèle hiérarchique. Or c'est à partir de ces variances que les fonctions de focalisation délimitent le sous-espace d'état des enfants d'un élément. Il en résulte qu'au fur et à mesure des détections réussies des primitives, la focalisation est de plus en plus fine et le système de perception de plus en plus efficace.

Finalement, il faut noter que, le parcours de toutes les branches utiles du réseau d'hypothèse

n'est pas forcément nécessaire à l'application : un système de perception pour être efficace n'a pas à être exhaustif. Un critère d'arrêt du processus récursif doit donc être défini.

4.3.4 Critère d'arrêt : évaluation de la vraisemblance et de la précision

Deux critères sont a priori importants pour évaluer la qualité de la perception au regard de la tâche robotique qui doit être accomplie. Il s'agit de la précision et de la vraisemblance.

L'estimation de l'état de la scène doit être précise pour que le robot puisse interagir avec elle de manière suffisamment fine. Comme cela est souligné dans la section précédente, la précision s'améliore après chaque détection réussie par la mise à jour de l'estimation de la covariance de l'état effectuée par l'EKF. L'amélioration de la précision dépend essentiellement de la variance du bruit d'observation de la primitive par son détecteur associé. Il est donc nécessaire de caractériser ce bruit.

Le processus de perception est réitéré jusqu'à ce que la précision requise soit atteinte.

La vraisemblance caractérise le niveau de confiance dans la suite d'hypothèse en cours de traitement. Le principe est simple, si une détection ne réussit pas, la confiance dans l'hypothèse diminue. La vraisemblance est mise à jour après chaque détection en même temps que l'état. L'estimation de la vraisemblance est effectuée dans le cadre d'un processus probabiliste (réseau bayésien dynamique).

De même, le processus de perception est réitéré jusqu'à ce que la vraisemblance requise soit atteinte.

4.4 Conclusion

Ce chapitre détaille le modèle de perception attentionnelle par focalisation progressive qui est le fil conducteur de mes travaux de recherche. Cette stratégie de perception reprend explicitement les remarques de R. Bajcsy mettant en avant la nécessité d'un processus de contrôle d'acquisition des données (ici de leur traitement bas niveau) en fonction de l'état courant de l'interprétation.

Elle inclut deux des processus de focalisation de l'attention proposés par Tsotsos :
– dans l'espace des caractéristiques par la définition de zones d'intérêt dans les données capteur,
– dans l'espace des décisions avec la sélection de l'objet à reconnaître dans la hiérarchie selon le meilleur compromis utilité/coût.

Il faut toutefois reconnaître qu'elle n'a pas pour origine une prise de conscience issue de la seule étude de la bibliographie de la nécessité de ce type d'approche attentionnelle pour la perception en robotique. Elle repose plutôt sur l'expérience acquise au fil du temps à la recherche de solutions efficaces pour la perception de l'environnement (routier en particulier). Cette réflexion a conduit à constater la capacité limitée des approches non actives à apporter ces solutions dans des cas très appliqués comme la détection des bords de voie pour l'aide à la conduite.

Elle propose finalement une solution originale pour relier les observations faites dans les données brutes des capteurs au modèle de la scène en utilisant explicitement ce modèle pour

contrôler l'observation. Elle tire partie de chaque observation au moment où elle est faite par la remise à jour du modèle (état et sa covariance). C'est sans doute ce dernier point, qui définit finalement le concept de focalisation progressive et qui constitue sa principale originalité. L'évaluation en ligne de la vraisemblance de l'hypothèse courante constitue également un point original. Cependant certains aspects comme la définition des fonctions d'utilité et de coût restent encore très dépendants des applications. Un travail fondamental reste à poursuivre sur ce point ainsi que sur la gestion des occultations ou sur l'apprentissage automatique du modèle de la scène qui nécessiterait d'ajouter un processus *bottom-up* au système attentionnel proposé.

Ce chapitre tente de formaliser ce système. Mais en robotique, formaliser ne suffit pas. Il faut nécessairement expérimenter. Le LASMEA est depuis de longues années reconnu pour la qualité de ces expérimentations dans le domaine de la robotique mobile en général et de la perception pour l'aide à la conduite en particulier. Pour ces deux raisons, l'impérieuse nécessité de confronter les concepts à la réalité s'est donc imposée d'elle même et les solutions développées ont été systématiquement testées sur des données réelles et si possible dans le cadre d'application embarquées. Ce souci me semble être une caractéristique importante de mon travail de recherche. Ainsi, les prochains chapitres présentent la mise en œuvre du système de perception proposé dans différentes applications d'aide à la conduite ou de robotique mobile auxquelles j'ai activement contribué depuis une dizaine d'années. Elles correspondent au travail réalisé dans le cadre de 5 thèses que j'ai co-encadrées (§3.2.2).

La description de chaque application est précédée d'une introduction destinée à en cadrer la portée par rapport au système de perception exposé ci-dessus sous une forme générale. Chacune de ces applications décline sous une forme adaptée ce même système de perception attentionnelle. Ainsi, l'ensemble du formalisme général présenté n'est pas mis en œuvre complètement dans chaque cas pour des raisons pratiques et chronologique : l'algorithme a évolué au cours du temps et dans ses versions précoces, il n'intégrait pas forcément l'ensemble des fonctionnalités. En particulier, l'aspect hiérarchique de l'analyse de scène ou d'objet complexe n'a jamais été mis en œuvre. En revanche, et c'est une contribution importante du travail présenté ici, le cas d'un système multisensoriel (caméra vidéo et télémètre LASER) est abordé dans deux applications.

5 Perception pour la localisation

5.1 Localisation d'un véhicule sur sa voie de circulation

5.1.1 Introduction

Le système de perception attentionnelle proposé a été initialement mis en œuvre pour apporter une solution au problème de la localisation d'un véhicule sur sa voie de circulation. Cette application correspond aux travaux de thèse de Romuald Aufrère soutenue en 2001 [72]. Dans cette toute première implémentation de l'algorithme, la perception est uniquement visuelle : l'analyse des images issues d'une caméra embarquée permet la détection des bords de voie dans les images. La localisation (position latérale par rapport à un bord et cap local) en est déduite. La scène est constituée d'un seul objet, la voie, formée de ses deux bords. L'estimation de la vraisemblance n'utilise pas un réseau bayésien mais un calcul de probabilité assez simple basé sur le dénombrement des détections réussies sur chacun des bords de voie.

Ici, le processus de perception n'est pas supervisé par une représentation particularisée de la scène : l'état est constitué d'un modèle générique de route dont la valeur courante est directement utilisée pour focaliser les détections.

L'application de localisation est donc déclinée dans un contexte local dans lequel il n'est pas possible de connaître a priori sur quelle section du réseau routier circule le véhicule puisqu'aucune information de positionnement globale n'est disponible. La carte de l'environnement qui guide la perception de la scène selon le processus attentionnel proposé est ici constituée d'un modèle géométrique générique de l'ensemble route/véhicule. Le véhicule circule *a priori* sur une voie de largeur L, de courbure C, à une distance x_0 du bord droit de la voie avec un angle de cap ψ et un tangage α. Ces cinq paramètres sont regroupés dans un vecteur d'état dont l'estimation est l'objectif du processus de perception. Une valeur initiale réaliste est affectée au vecteur d'état et à sa matrice de covariance, ce qui constitue pour cette application, l'étape d'apprentissage du système de perception. Ces paramètres sont reliés par une projection perspective à la position des bords de la voie dans l'image. Les bords sont représentés par un nombre fini de segments de droite qui sont ainsi reliés statistiquement au vecteur d'état et à sa matrice de covariance. C'est ce lien statistique qui sous-tend le processus attentionnel conformément au principe de focalisation progressive présenté section 4.3. L'intervalle de confiance délimité dans l'espace d'état par la matrice de covariance se projette dans l'espace image. Une liste de zones d'intérêt de l'image peuvent ainsi être définies. Un détecteur (opérateur de traitement d'image) est appliqué dans la zone d'intérêt où un bord de voie a la probabilité la plus importante de se trouver. Cette détection peut échouer. Dans ce cas, le processus de détection est appliqué dans la zone d'intérêt suivante de la liste. Si la détection réussit, son résultat est utilisé pour remettre à jour le vecteur d'état et une nouvelle liste de zones d'intérêt est générée. Ce processus (sélection de zone d'intérêt, détection, remise à jour) est réitéré jusqu'à ce qu'un critère de reconnaissance de la route soit atteint.

La mise en œuvre du processus attentionnel de focalisation de la détection des bords dans des zones d'intérêt apporte une contribution intéressante en montrant qu'une analyse exhaustive de l'image n'est pas nécessaire pour extraire l'information utile. La réduction considérable du temps de calcul, la robustesse apportée par l'amélioration du rapport signal sur bruit ainsi que l'intégration dans un même processus, de la reconnaissance de la route, du suivi des bords et de la localisation route/véhicule ont fait le succès de cette méthode.

5.2 Description détaillée de l'application

5.2.1 Motivations de l'application et état de l'art en détection de bord de voies

La communauté scientifique est depuis longtemps très active pour le développement d'aides à la conduite. Plusieurs systèmes ont été développés dans ce domaine par des équipes de recherche : MOB-LAB[2], RALPH[23], VAMORS[22], ou dans le cadre de projets de recherche (PROMETHEUS, ARCOS, etc...)

Le travail présenté dans cette partie porte sur la reconnaissance de bords de route et la localisation du véhicule sur sa voie qui sont des composantes majeures de tels projets. La plupart des approches permettant cette reconnaissance à la fois sur routes marquées et non marquées sont basées sur l'exploitation des bords de la route dans l'image. En effet dans le contexte autoroutier, une segmentation en région de l'image est délicate en raison du trafic et de la discontinuité du marquage. Il faut noter que dans le cas des routes non marquées, c'est au contraire ce type de segmentation qui s'avère le plus efficace.

Dans ce domaine de la reconnaissance de bords de route, deux catégories d'approches sont principalement utilisées : les approches image et les approches 4D.

La première exploite un modèle de la route dans l'image afin de grouper les indices préalablement détectés et de remettre à jour le modèle dans l'image. Certaines approches reposent sur une modélisation rectiligne de la route comme Chen [48]. Lakshmanan dans [49] utilise des hyperboles dont la mise à jour est réalisée à l'aide d'indices issus d'une analyse fréquentielle. Thorpe et Pommerleau [23] ont développé le système RALPH basé sur la minimisation des paramètres d'une courbe par rapport à une reprojection vue de dessus de la route. Broggi [2] exploite une représentation similaire associée à une approche morphologique. D'autres approches utilisent des modèles plus complexes comme par exemple celle de Guichard [41] qui effectue la recherche des indices selon les courbes cohérentes les plus longues dans l'image.

L'inconvénient majeur de cette approche image, excepté dans le cas rectiligne, est que la recherche de la cohérence entre les bords de la route dans l'image est difficile. Cette cohérence permet par exemple de détecter avec précision les bords partiellement occultés en utilisant la connaissance de l'autre bord.

La seconde catégorie, celle des approches 4D consiste à prendre en compte la connaissance dynamique du véhicule et son évolution sur la route. La cohérence des bords de la route est ainsi assurée.

Le principe est de modéliser la route ainsi que le véhicule et son évolution dynamique dans l'espace 3D. Les mesures faites dans l'image permettent de mettre à jour les paramètres du modèle et ceux de l'état du véhicule. Cette approche présente l'avantage du réalisme des modèles utilisés qui peuvent être complexes comme celui de Dickmanns [22]. Elle permet de

prédire de manière précise la position des primitives pour l'image suivante.

Une phase d'initialisation est cependant nécessaire et elle peut représenter un inconvénient en cas de perte de suivi car la durée de ré-initialisation peut être importante et les conditions initiales non respectées. Cette approche 4D est beaucoup moins stable et plus sensible au bruit qu'une approche image en raison de l'approximation des petits angles, des erreurs d'étalonnage et des vibrations du véhicule.

Compte tenu de ces différentes constats l'approche choisie est de type image. Elle gère la cohérence entre les bords de la route évitant ainsi le principal écueil de ce type de méthode. Bien entendu le développement d'une telle approche nécessite ensuite l'utilisation d'un modèle caractérisant l'ensemble route/véhicule afin de localiser le véhicule sur sa voie de circulation (objectif final de l'application). De plus, la solution développée doit être rapide, efficace vis-à-vis des occultations et adaptable à tous types de routes.

Les méthodes traditionnelles pour étudier ce type d'approche utilisent des techniques de reconnaissance des formes. Les approches travaillant par accumulation dans l'espace des transformations (transformée de Hough [13]) sont adaptées à tous les problèmes où la dimension du vecteur de paramètres est faible (≤ 6). Ici un modèle d'ordre élevé est choisi afin de modéliser avec précision les bords de la route. Les approches développées dans l'espace des correspondances permettent de travailler dans des espaces d'ordre élevé mais font généralement apparaître un problème de gestion de la combinatoire. Dans ce cas, chaque détection est associée tour à tour à chacune des primitives formant l'objet à reconnaître. Cette technique ne nécessite pas une détection globale des primitives dans l'image comme pour les méthodes d'accumulation.

La solution proposée ici est constituée de quatre modules (figure 5.1) : reconnaissance, localisation, suivi et gestion du nombre de voies.

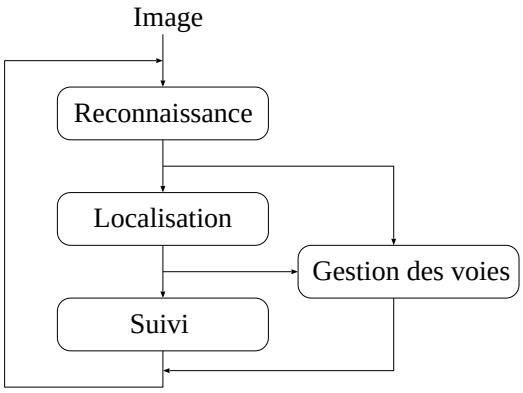


FIG. 5.1 – Organigramme de l'application de localisation du véhicule sur sa voie.

Les images sont fournies par une caméra embarquée au niveau du rétroviseur intérieur du véhicule. L'étape de reconnaissance permet de déterminer précisément la position de la route dans l'image.

Elle est basée sur l'exploitation d'un modèle statistique des bords de la voie de circulation. La méthode peut être appliquée à la fois sur des images de routes marquées et non marquées.

A partir du résultat de l'étape de reconnaissance, la phase de localisation permet de calculer la position latérale et l'orientation du véhicule sur sa voie. Cette étape permet de déterminer également des paramètres relatifs à la caméra (angle d'inclinaison) et à la route (largeur et

courbure latérale).

Le module de suivi est alors appliqué afin de limiter l'espace de recherche des bords de la route dans l'image suivante.

Le dernier module gère le nombre de voies de circulation (sur autoroutes) et identifie le numéro de la voie de circulation du véhicule.

5.2.2 Reconnaissance

Cette application est basée sur la reconnaissance récursive des bords de la route dans l'image guidée par un modèle probabiliste générique de l'ensemble route/véhicule (figure 5.2).

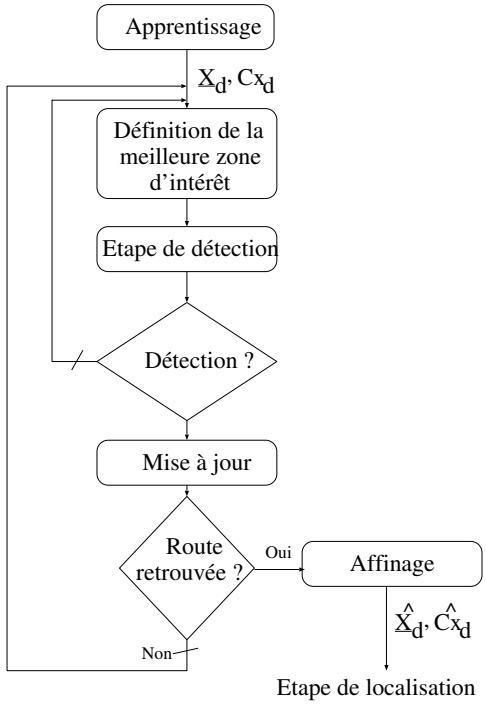


FIG. 5.2 – Organigramme de l'étape de reconnaissance.

L'organigramme de la figure 5.2 fait apparaître le principe de perception attentionnelle par focalisation progressive mis en œuvre pour le première fois dans cette application [3].

5.2.3 Modèle

Le modèle est composé d'un vecteur $\underline{X}_d$ et de sa matrice de covariance $\mathbf{C}_{X_d}$ définis par :

$$\underline{X}_d = \begin{pmatrix} u_{1g} \\ . \\ u_{ng} \\ u_{1d} \\ . \\ u_{nd} \end{pmatrix}, \mathbf{C}_{X_d} = \begin{pmatrix} \sigma^2_{u_{1g}} & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & \sigma^2_{u_{ng}} & . & . & . \\ . & . & . & \sigma^2_{u_{1d}} & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & \sigma^2_{u_{nd}} \end{pmatrix} \tag{5.1}$$

Les éléments du vecteur d'état précédemment notées x_i dans la section 4.3.1 correspondent ici représentent aux abscisses $(u_{1g},...,u_{ng},u_{1d},...,u_{nd})^t$ dans le repère de l'image de chaque point des deux bords de la route et pour différentes ordonnées (lignes de l'image) v_i fixées une fois pour toutes ($i \in [1,n]$) (typiquement $n = 10$). La valeur de n peut varier suivant l'application (autoroutes, routes nationales, routes de campagne, etc.). La matrice $\mathbf{C}_{X_d}$ définit les variations possibles de u_{ig} et u_{id} dans l'image.

Les bords de la route dans l'image sont représentés un ensemble de $2n-2$ segments ($n-1$ pour le bord gauche et $n-1$ pour le bord droit) caractérisés par les coordonnées de leurs extrémités $(u_i,v_i)^t$ et $(u_{i+1},v_{i+1})^t$. Les abscisses u_i et u_{i+1} sont les paramètres contenus dans le vecteur modèle $\underline{X_d}$ pour les ordonnées v_i et v_{i+1} définies précédemment. Exceptés les points extrêmes du vecteur modèle u_1 et u_n, chaque paramètre appartient donc à deux segments consécutifs.

La valeur initiale du modèle ($\underline{X_d}$, $\mathbf{C}_{X_d}$) est obtenue par une phase d'apprentissage. Cette phase utilise un modèle développé par Roland Chapuis dans sa thèse [68] qui exprime les coordonnées u_i en fonction de $(v_i,z_0,x_0,\psi,\alpha,C,L)$ où z_0 est la hauteur de la caméra, x_0 la position latérale du véhicule sur la chaussée par rapport à l'un des bords de voie, ψ l'angle d'orientation (cap) du véhicule, α l'angle d'inclinaison verticale de la caméra, C la courbure latérale de la route et L sa largeur (voir équations (5.2) et (5.3)).
Cette modélisation est utilisée uniquement pour la phase d'apprentissage et n'intervient donc pas dans la procédure de reconnaissance.

$$u_{ig} = e_u\{\frac{e_v z_0}{2(v_i - e_v\alpha)}C_l - \frac{v_i - e_v\alpha}{e_v z_0}(x_0 - L) + \psi\} \tag{5.2}$$

$$u_{id} = e_u\{\frac{e_v z_0}{2(v_i - e_v\alpha)}C_l - \frac{v_i - e_v\alpha}{e_v z_0}x_0 + \psi\} \tag{5.3}$$

e_u et e_v sont les paramètres intrinsèques de la caméra (supposée calibrée).

Ainsi, le vecteur moyen $\underline{X_d}$ peut être facilement obtenu à partir d'une connaissance moyenne des différentes variables définissant le véhicule, la route et la caméra (x_0,L,ψ,C_l,α). La matrice de covariance $\mathbf{C}_{X_d}$ est calculée en utilisant la jacobienne des équations (5.2) et (5.3) et une matrice de covariance définissant les variations possibles des paramètres moyens précédemment définis suivant le contexte envisagé.
Cette phase est nécessaire afin de rechercher seulement les configurations de route réalistes dans l'image. La figure 5.3 montre le résultat de la phase d'apprentissage sur une image donnée.

Pour chaque bord de la voie, un intervalle de confiance centré sur $\underline{X_d}$ est représenté (courbes blanches) dont la largeur est déduite de la matrice $\mathbf{C}_{X_d}$. Les ordonnées v_i sont également représentées (lignes noires). Elles correspondent aux hauteurs des zones de recherche. Le nombre total de zones de recherche à tester dans ce cas est 18.

Procédure de recherche

Dans cette procédure, le but est de rechercher la valeur optimale de $\underline{X_d}$ représentant les bords de la route dans l'image en cours d'analyse.

La figure 5.4 présente un exemple d'évolution de l'algorithme pour une image donnée.

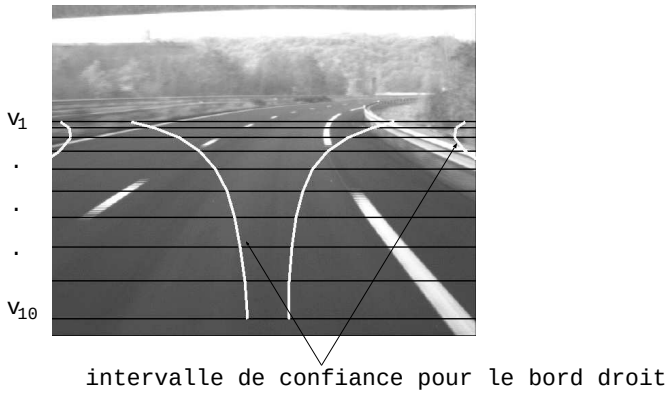


FIG. 5.3 – Modéle initial résultat de la phase d'apprentissage.

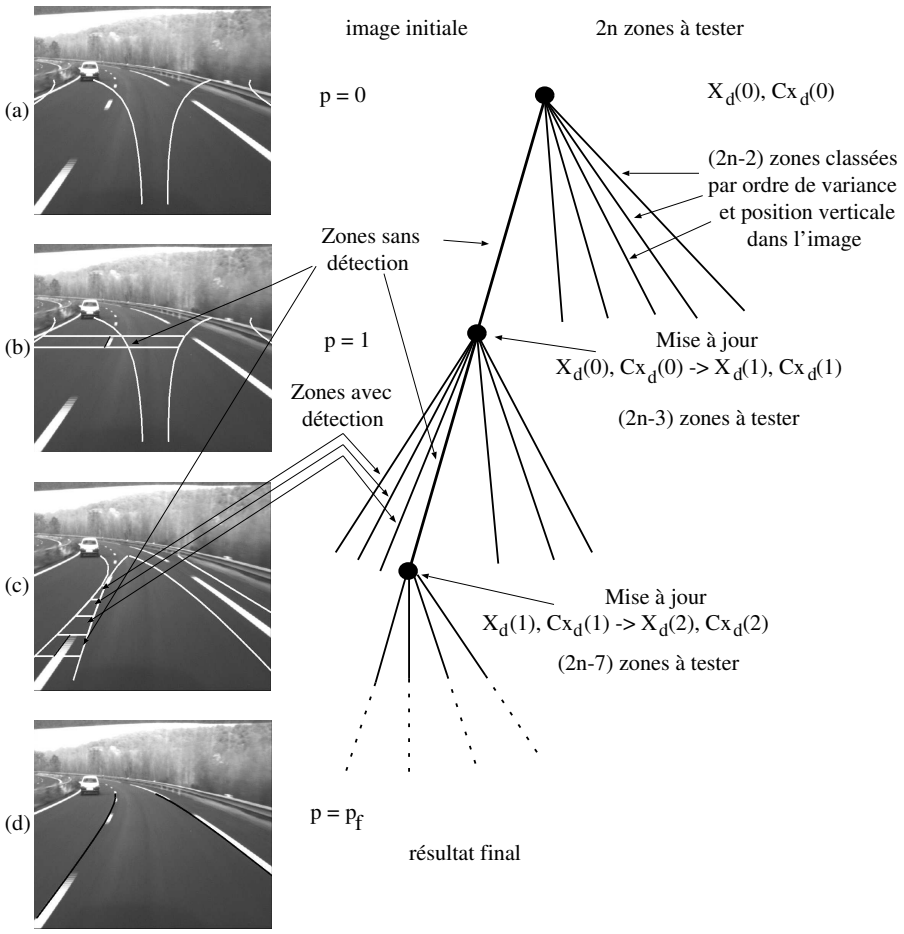


FIG. 5.4 – Evolution de l'algorithme sur une image

Fonction de coût

La fonction de coût utilisée ici pour la sélection de l'observation à effectuer, est simplement la variance des composantes du vecteur d'état. Après la phase d'apprentissage (profondeur d'analyse $p = 0$, figure 5.4-a), la plus petite variance de la diagonale de $\mathbf{C}_{X_d}$ est retenue car elle permet de définir la zone d'intérêt (voir paragraphe suivant) de taille la plus faible conduisant ainsi à une durée minimale de la détection. Le critère de coût correspond donc dans ce cas

au temps de calcul du traitement d'image bas niveau. De plus, ce choix correspond *a priori* au bord dont la présence est la plus probable.

Fonction de focalisation et définition de la zone d'intérêt

La variance $\sigma^2_{u_i}$ qui correspond à la variance la plus faible de $\mathbf{C}_{X_d}$ (équation 5.1) permet de construire une zone d'intérêt. Cette zone peut être caractérisée par un trapèze dont les sommets sont définis de la façon suivante :

$P_1 = (u_i - \sigma_{u_i}, v_i)$,
$P_2 = (u_i + \sigma_{u_i}, v_i)$,
$P_3 = (u_{i+1} - \sigma_{u_{i+1}}, v_{i+1})$,
$P_4 = (u_{i+1} + \sigma_{u_{i+1}}, v_{i+1})$

Détecteur

Pour chaque ligne de l'image appartenant à cette zone, le point de gradient maximum est localisé. A l'ensemble des points ainsi constitué, un segment de droite est ajusté par la méthode des moindres carrés médians (figure 5.4-b). Cette méthode tient compte d'une contrainte sur la pente du segment à détecter (déduite des variations des paramètres u_i et u_{i+1} permettant de définir la zone d'intérêt donc de $\mathbf{C}_{X_d}$) afin de fournir un segment compatible avec le modèle courant. Ceci est une déclinaison de la focalisation dans l'espace des tâches : les seuils du détecteur sont adaptés automatiquement en fonction des nécessités de la perception.

Il se peut qu'aucun segment ne soit détecté dans la zone considérée (figure 5.4-c). Dans ce cas, une détection est tentée dans une autre zone d'intérêt définie par la plus petite variance suivante de $\mathbf{C}_{X_d}$.

Si aucun segment n'est détecté dans l'ensemble des zones non testées $(2N - p)$ à la profondeur d'itération donnée p, l'algorithme abandonne cette hypothèse et revient à la profondeur $p-1$ et réitère le processus sur une nouvelle hypothèse.

Une détection fournit deux points (intersections entre le segment détecté et les bords de la zone de recherche) qui sont utilisés par un filtre de Kalman "dégénéré" (aucune évolution du vecteur d'état) pour mettre à jour $\underline{X_d}$ et $\mathbf{C}_{X_d}$ (figure 5.4-c).

Vraisemblance de l'hypothèse courante - Critère d'arrêt du processus récursif

Le système de perception (choix d'une zone d'intérêt, détection, mise à jour) est réitéré et s'arrête lorsqu'un critère Q est atteint. Il est défini par dénombrement des détections réussies sur chacun des bords (équation (5.4)) :

$$Q = \{P_g = \frac{N_g}{\Delta_v} > 5\% \text{ et } P_d = \frac{N_d}{\Delta_v} > 5\% \text{ et } P_t = \frac{N_g + N_d}{2\Delta_v} > 25\%\} \tag{5.4}$$

où
N_g et N_d sont respectivement le nombre de points détectés (hypothèses confirmées) sur les bords gauche et droit de la route

et
Δ_v (environ 300 pixels) représente la hauteur de la zone d'analyse ($v_n - v_1$).

Les seuils sur les éléments P_g P_d et P_t du critère Q ont été fixés selon des considérations expérimentales.

Le critère Q peut être atteint sans une analyse complète de l'hypothèse courante. Par conséquent, toutes les zones n'ont pas été testées.

Dans ce cas, afin de fournir une précision optimale, des détections sont tentées dans les zones encore non traitées. Le résultat final correspond au vecteur $\underline{X_d}$ ayant conduit à la plus grande probabilité P_t (figure 5.4-d).

Si Q n'est pas atteint, l'algorithme considère que la route n'est pas retrouvée dans l'image mais il fournit malgré tout la meilleure solution.

Le vecteur $\underline{X_d}$ ayant conduit au résultat de la phase de reconnaissance pour l'image k est noté $\underline{\widehat{X}_d}(k)$.

5.2.4 Localisation et suivi

L'objectif est de calculer la position et l'orientation du véhicule sur sa voie à partir du résultat de la reconnaissance $\underline{\widehat{X}_d}(k)$ et de fournir un nouvel espace de recherche des bords dans l'image suivante $k+1$.

Localisation

Pour cette étape, le modèle est complété : $\underline{X} = (\underline{X_d}, \underline{X_l})^t$. Il contient le vecteur $\underline{X_d}$ précédemment défini et un vecteur $\underline{X_l} = (x_0, \psi, \alpha, C_l, L)^t$ qui regroupe les paramètres de localisation.

Le nouveau modèle défini par $\underline{X}$ et $\mathbf{C}_X$ est alors :

$$\underline{X} = \begin{pmatrix} \underline{X_d} \\ \underline{X_l} \end{pmatrix} \text{ et } \mathbf{C}_X = \begin{pmatrix} \mathbf{C}_{X_d} & \vdots \\ \cdots & \mathbf{C}_{X_l} \end{pmatrix} \tag{5.5}$$

$\mathbf{C}_{X_l}$ est la matrice de covariance de $\underline{X_l}$. La valeur initiale de ce modèle $(\underline{X}, \mathbf{C}_X)$ est obtenue par une phase d'apprentissage similaire à celle utilisée pour définir ($\underline{X_d}$, $\mathbf{C}_{X_d}$) car les équations (5.2) et (5.3) expriment une relation entre les différents paramètres constituant $\underline{X}$.
La procédure de reconnaissance fournit, pour l'image k, un vecteur d'observation $\underline{\widehat{X}_d}(k)$, composé des différents paramètres définissant les bords de la route dans l'image (u_{ig} et u_{id}) et sa matrice de covariance $\widehat{\mathbf{C}}_{X_d}(k)$.
La mise à jour du vecteur $\underline{X}$ et de sa covariance $\mathbf{C}_X$, pour une image k, est effectuée par un filtre de Kalman :

$$\begin{cases} \underline{X}(k) = \underline{X}(k-1) + \mathbf{K}[\underline{\widehat{X}_d}(k) - \mathbf{H}\underline{X}(k-1)] \\ \mathbf{C}_X(k) = \mathbf{C}_X(k-1) - \mathbf{K}\mathbf{H}\mathbf{C}_X(k-1) \end{cases} \tag{5.6}$$

– $\mathbf{K} = \mathbf{C}_X(k-1)\mathbf{H}^t\left[\mathbf{H}\mathbf{C}_X(k-1)\mathbf{H}^t + \mathbf{C}_{X_d}(k)\right]^{-1}$
– $\mathbf{H}$ telle que $\underline{\widehat{X}_d}(k) = \mathbf{H}\underline{X}(k) + \underline{w}$,

– $\underline{\widehat{X}}_d$: vecteur résultat du processus de détection,
– $\mathbf{C}_{X_d} = E[\underline{w}\underline{w}^t]$: matrice de covariance $\underline{\widehat{X}}_d$.
$\underline{X}(k)$ fournit une estimation du vecteur $\underline{X}_l(k)$ pour l'image k. Ce vecteur permet ainsi de caractériser la position et l'orientation du véhicule sur sa voie ainsi que la courbure et la largeur de la route et l'angle d'inclinaison de la caméra. L'estimation de cet angle permet de prendre en considération les phénomènes de tangage contrairement à la plupart des approches existantes qui considèrent cet angle comme constant.

Suivi

Le suivi consiste à faire évoluer le modèle des bords obtenu à l'image k afin de focaliser l'attention du module de reconnaissance pour l'image $k+1$. Le modèle à la profondeur $p = 0$ utilisé pour cette image n'est donc pas celui fourni par la phase d'apprentissage qui ne doit donc être a priori utilisé que pour la toute première image.

Le problème revient à calculer $\underline{X}(k+1)$ et $\mathbf{C}_X(k+1)$ à partir du modèle à l'image k : $\underline{X}(k)$, $\mathbf{C}_X(k)$ et du déplacement du véhicule entre deux images successives.

Pour réaliser ce processus, un filtre de Kalman prend en compte ces différents paramètres et les erreurs liées à l'évolution entre deux images.
Grâce à cette évolution le vecteur $\underline{X}_d(k+1)$ et sa matrice de covariance $\mathbf{C}_{X_d}(k+1)$, inclus dans $\underline{X}(k+1)$ et $\mathbf{C}_X(k+1)$, permettent de définir l'espace de recherche pour l'image suivante.

La figure 5.5(a) montre la zone de recherche initiale des bandes dans l'image (lignes noires pour le bord gauche et blanches pour le bord droit) issue de la phase d'apprentissage et la figure 5.5(b) présente la zone initiale en phase de suivi. L'espace de recherche est considérablement réduit comme, par conséquent, les temps de calcul de l'étape de reconnaissance.

(a) (b)

FIG. 5.5 – Zone de recherche initiale (a) sans suivi et (b) avec suivi.

Pour gérer le changement de voies, le modèle $(\underline{X}, \mathbf{C}_X)$ est mis à jour, permettant ainsi de définir la nouvelle zone de recherche, avec une valeur de x_0 adaptée à la configuration donnée.

$$\begin{cases} Si \quad x_0(k) > L \Rightarrow x_0(k+1) = x_0(k) - L \\ Si \quad x_0(k) < 0 \Rightarrow x_0(k+1) = x_0(k) + L \end{cases} \tag{5.7}$$

5.2.5 Gestion du nombre de voies

L'algorithme est capable de gérer le nombre de voies. Pour réaliser ce processus, $2n$ paramètres sont ajoutés au vecteur $\underline{X}_d$ afin de détecter à la fois les bords de la voie de circulation et deux lignes latérales. Ce nouveau modèle ($4n$ paramètres image) est également initialisé

lors d'une phase d'apprentissage.
A partir du résultat de l'étape de reconnaissance (1 voie = $2n$ paramètres), une mise à jour du modèle complet ($4n$ paramètres) permet de limiter l'espace de recherche des bandes latérales pour une même image. Ensuite, le processus de reconnaissance est redémarré dans les $(2n-2)$ zones correspondantes aux bandes latérales (($n-1$) pour la bande de gauche et ($n-1$) pour la bande de droite).

Une bande latérale est considérée comme reconnue si la probabilité associée ($\frac{N_{gg}}{\Delta_v}$ et/ou $\frac{N_{dd}}{\Delta_v}$) est plus grande que 5%. N_{gg} et N_{dd} représentent les nombres de points détectés sur chacune des bandes latérales. La détection d'une bande permet de mettre à jour une probabilité définissant la présence de voies latérales. Ensuite, si une ou deux bandes latérales sont considérées comme retrouvées, l'étape de reconnaissance est redémarrée, pour l'image suivante, sur deux ou trois voies ($3n$ ou $4n$ paramètres image). Les probabilités sur la présence de bandes latérales sont réinitialisées toutes les 80 images afin de considérer un changement éventuel de configuration de l'autoroute (passage de trois à deux voies de circulation).
Une probabilité est également associée à chacune des voies de circulation détectée par le processus. L'estimation de la position latérale du véhicule permet d'augmenter ou de diminuer la probabilité de présence du véhicule dans chacune des bandes. Le processus de changement de voie permet de faire basculer ces différentes probabilités.

5.2.6 Résultats

Afin de valider l'approche dans un contexte temps réel, l'algorithme a été implanté sur une architecture embarquée. La figure 5.15 présente des résultats relatifs à l'étape de reconnaissance à la fois sur routes marquées et non marquées. La caméra utilisée a une focale $f = 12mm$ et elle est placée à une hauteur $z_0 = 1,5m$. Les voies sont parfaitement détectées malgré les ombres et les surbrillances. Pour la dernière ligne de résultats de la figure 5.6, la route n'a pas été reconnue car le critère Q (voir § 5.2.3) n'a pas été atteint mais la meilleure solution est fournie (P_t maximum).

FIG. 5.6 – Résultats sur routes marquées et non marquées.

La figure 5.7 montre des images où plusieurs voies de circulation sont présentes.

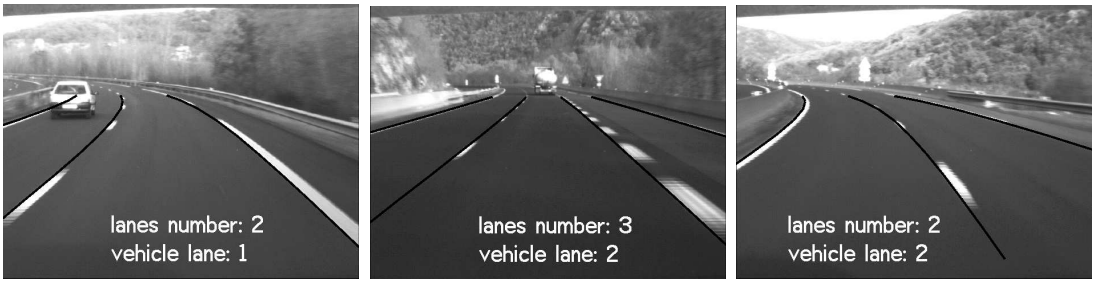


FIG. 5.7 – Résultats du module de gestion du nombre de voies.

L'algorithme considère la route comme retrouvée dans l'image pour un nombre d'itérations (nombre d'hypothèses testées) compris entre 18 et 150. Lorsque l'algorithme considère la route comme non retrouvée dans l'image, la meilleure solution en phase de suivi est obtenue avant 200 itérations, dans la majorité des cas.

Conclusion

Ce chapitre décrit une méthode de détection et de suivi de route qui peut être utilisée pour l'assistance à la conduite. Les principaux avantages de cette méthode sont la précision (ordre élevé du modèle), le principe de détection locale (mais la recherche est globale), la faible sensibilité aux occultations ou aux imperfections du marquage et la possibilité d'appliquer la méthode sur routes marquées et non marquées. De plus, cette approche est capable de gérer le nombre de voies.

Grâce à cette première mise en œuvre du système de perception attentionnelle proposé, cette méthode s'est démarquée de ses "concurrentes". En effet la focalisation progressive permet de considérer que les différentes primitives à détecter dans l'image pour reconnaître la route sont statistiquement liées entre elles. Ce lien est explicitement : en rendant les détections dépendantes, elle assure la cohérence de la reconnaissance : l'objet (i.e. la route) n'est pas considéré comme une série de primitives juxtaposées mais comme un ensemble.

La limitation *a priori* de l'espace d'état aux seules configurations réalistes assure que la méthode ne reconnaisse que des routes compatibles avec ces configurations. Ceci permet par exemple d'exclure ou de limiter fortement l'influence de données qui, pour être tout à fait déterministes, n'en sont pas moins "indésirables" comme par exemple les bords de voie d'accélération ou de sortie d'autoroute.

Il faut aussi noter que dans cette première implémentation, la reconnaissance par focalisation progressive n'inclut pas de fonction d'utilité. Ceci n'est pas un réel problème dans la mesure où chaque primitive concourt à son tour à la reconnaissance des bords de voie sans pour autant que l'une d'entre elles soit plus utile qu'une autre pour atteindre ce but. En revanche le modèle statistique gaussien qui est utilisé leur confère un coût de détection différent qui lui, est exploité.

Cette méthode a été valorisée par 3 articles, 8 communications, 6 collaborations industrielles ou contrats de recherche ainsi que dans le cadre du projet ARCOS du PREDIT.

Au bilan, cette première application a permis de valider sur un exemple le concept de la perception attentionnelle par focalisation progressive.

5.3 Localisation d'un véhicule dans un environnement cartographié

5.3.1 Introduction

La méthode précédente est utilisée comme "brique vision" dans un processus de fusion de données multi-capteurs destiné à la localisation globale du véhicule. Cette application correspond aux travaux menés dans le cadre de la thèse de Jean Laneurit soutenue le 3 juillet 2006 [45]. Couplée à une carte numérique précise de la section routière où circule le véhicule, elle permet de calculer sa pose dans le repère absolu de la carte. Les paramètres de localisation ainsi calculés sont fusionnés avec les informations issues d'un GPS peu précis et de capteurs proprioceptifs qui servent à identifier les paramètres d'un modèle cinématique du véhicule. La fusion d'informations conduit à une localisation globale du véhicule avec une précision métrique tout en étant peu sensible aux occultations des satellites (source de perte d'information). Ce travail a de plus été l'occasion d'une modélisation des erreurs sur les informations GPS qui sont non blanches. La fusion d'informations permet l'estimation du biais qui peut alors être soustrait ce qui améliore la précision de la localisation par GPS.

Cette application utilise de manière conjointe l'approche précédente de détection suivi de bord de voies (§ 5.1) dont le principe repose sur le système de perception proposé et une représentation de la scène sous la forme d'une carte d'amers géoréférencés. Elle constitue de ce fait une étape importante vers une localisation absolue qui exploite explicitement cette carte en temps que modèle du monde pour guider le processus de perception attentionnelle des amers.

5.3.2 Description détaillée de l'application

Intérêt de l'application et éléments de l'état de l'art en localisation

Des systèmes d'aide à la navigation d'un véhicule sur le réseau routier sont de plus en plus souvent proposés aux automobilistes. Ils permettent la localisation avec une précision de l'ordre de 10m qui suffit à une navigation correcte même dans les milieux urbains.

L'évolution actuelle des systèmes d'aide à la conduite s'oriente vers le développement de fonction visant à l'amélioration de la sécurité. Il ne s'agit plus d'une simple aide passive à la navigation mais de solutions actives pour lesquelles une localisation avec une précision décamétrique s'avère insuffisante. Les systèmes de localisation pour la navigation sont donc inadaptés.

Les nombreuses méthodes déjà développées pour la localisation d'un véhicule ou d'un robot dans son environnement sont classées en trois catégories : localisation relative, absolue et hybride [43].

Localisation relative La localisation relative (à l'estime) fournit la position et l'orientation d'un véhicule grâce à la connaissance de ses déplacements successifs depuis une configuration de départ connue. Des capteurs proprioceptifs permettent la mesure de grandeurs cinématiques propres au comportement du véhicule : distance parcourue, vitesse, accélération ou angle de giration et angle d'attitude [81], [7]. L'estimation de la position du véhicule est actualisée avec

une fréquence élevée. La localisation relative est une technique de positionnement très simple à mettre en œuvre. Le principal inconvénient est la dérive de l'estimation de la position du véhicule dans le temps.

Localisation absolue La localisation absolue donne la position du véhicule ou robot dans son milieu d'évolution. Elle repose sur des capteurs extéroceptifs et sur des points de repères naturels [87] ou artificiels (satellites GPS ou balises radiofréquences terrestres) [82], [9], [11]. Une connaissance *a priori* est requise sous la forme d'une carte de l'environnement [78]. Cette technique n'occasionne pas de dérive dans le temps. Son principal inconvénient est la perte de visibilité des points de repère qui peut entraîner l'impossibilité de fournir une position intègre c'est-à-dire pour laquelle le processus est susceptible de garantir que l'intervalle de confiance sur l'estimée de la position contient effectivement la position "vraie".

Localisation hybride Un comparatif des avantages et inconvénients des techniques de localisation absolue et relative fait apparaître leur complémentarité (tableau 5.1). De nombreux

Localisation relative	**Localisation absolue**
(-) Accumulation d'erreurs avec la distance	(+) Erreurs indépendantes de la distance
(-) Repère attaché à la position initiale du véhicule	(+) Repère rattaché à l'environnement
(+) Autonomie vis-à-vis de l'environnement	(-) Aménagement parfois nécessaire
(+) Fonctionnement à fréquence élevée	(-) Fonctionnement à fréquence très faible (-) Pertes d'informations possible

TAB. 5.1 – Comparatif localisation relative - localisation absolue.

systèmes de localisation combinent ces deux techniques, afin de pallier leurs défauts et de cumuler leurs avantages. La fusion de données (ou fusion multicapteurs) est réalisée par un filtre de Kalman étendu [9], par un UKF[1] [20] ou par un filtre à particule [12].

L'objectif est ici d'estimer l'attitude d'un véhicule dans un repère bidimensionnel du monde. Cette attitude est représentée par le vecteur d'état $\underline{X} = (x, y, \gamma, b_x, b_y)^t$ avec, x et y les coordonnées de la position du véhicule dans le monde, γ orientation du véhicule autour de l'axe vertical, b_x et b_y les biais sur les axes x et y des mesures GPS. La nécessité d'inclure ces deux termes dans le vecteur d'état est explicitée ci-après (§ 5.3.2).

La matrice de covariance $\mathbf{Q}$ du vecteur d'état permet de connaître la confiance sur l'estimation de chacun des paramètres. Elle s'écrit :

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_a & \mathbf{Q}_{ab} \\ \mathbf{Q}_{ba} & \mathbf{Q}_b \end{pmatrix}$$

avec $\mathbf{Q}_a$ la matrice de covariance associée à l'attitude du véhicule $(x, y, \gamma)^t$ et $\mathbf{Q}_b$ la matrice de covariance associée au biais $(b_x, b_y)^t$.

[1] Unscented Kalman Filter

Le véhicule expérimental est équipé de capteurs : un odomètre, un gyromètre, une caméra et un GPS bas coût, ainsi que d'une carte numérique précise de l'environnement. La localisation est réalisée par une méthode hybride qui combine les informations des capteurs proprioceptifs et extéroceptifs. La fusion des données est effectuée par un filtre de Kalman.

Trois types d'information sont fusionnés dans l'approche proposée : localisation absolue par GPS, localisation absolue par coopération vision/cartographie, localisation relative par un gyromètre un odomètre et angle au volant. Ils sont détaillés ci-après.

Localisation par GPS

Un GPS bas coût est utilisé. Il fournit les coordonnées de la position du véhicule dans le référentiel géocentrique Lambert 2.

Identification des erreurs Pour identifier les erreurs, les données GPS d'un même point ont été acquises pendant plusieurs heures sans déplacer l'antenne de réception. Les erreurs du GPS ne sont pas blanches. Elles peuvent être modélisées par un biais $\underline{b} = (b_x, b_y)^t$ qui évolue lentement dans le temps additionné d'un bruit $\underline{v}$ blanc et gaussien de faible amplitude. A chaque donnée GPS, le biais peut varier brutalement. La visibilité des satellites et les multtrajets en sont les principales origines. A cause du biais, les erreurs ne sont pas blanches comme le montrent les fonctions d'autocorrélation de b_x et de b_y (figure 5.8). La réactualisation du vecteur d'état est alors délicate. Le biais $\underline{b}$ et le bruit sur l'observation $\underline{v}$ sont séparés : les deux composantes du biais sont intégrées au vecteur d'état ce qui permet de garder un bruit $\underline{v}$ à spectre blanc compatible avec une estimation par filtre de Kalman. Le biais est donc estimé en même temps que la position. Ceci est rendu possible en raison forte corrélation du biais avec sa valeur précédente.

Précision du GPS Deux facteurs interviennent directement dans le calcul de la précision : le positionnement des satellites par rapport au récepteur qui est représenté par la valeur GDOP (Geometric Dilution of Precision) et la précision de la mesure de la distance entre l'utilisateur et chaque satellite UERE (User Equivalent Ranging Error) qui est de l'ordre de $4,2m$.

La précision σ_P peut alors s'exprimer par le produit de l'erreur de distance équivalente et du facteur représentatif de la qualité de la géométrie satellites-utilisateur :

$$\sigma_P = GDOP \times UERE$$

La figure 5.9 représente le modèle d'erreur lié au GPS.

Modèle d'observation

Le modèle d'observation pour les données GPS est de la forme :

$$\underline{X}_{gps} = \underline{X}_k + \underline{b}_k + \underline{v}_k$$

avec $\underline{X}_{gps} = (x_{gps}, y_{gps})^t$ les coordonnées du véhicule fournies par le GPS, $\underline{b} = (b_x, b_y)^t$ le biais sur les données GPS et $\underline{v}$ le vecteur de bruit sur les données GPS.

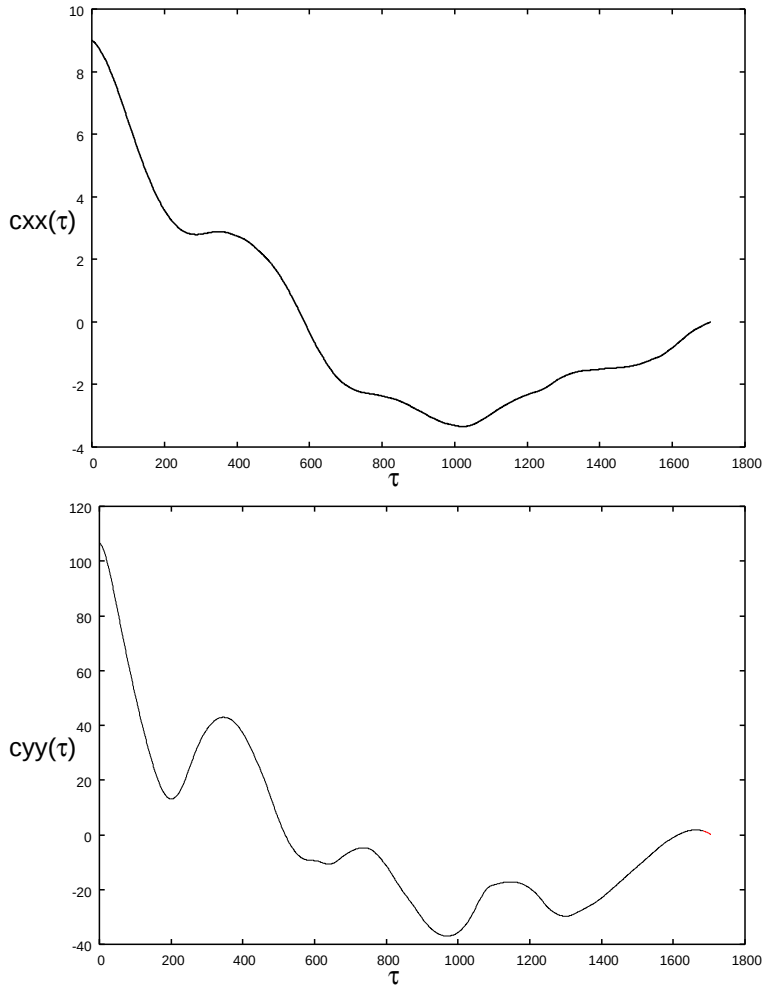


FIG. 5.8 – Fonction d'autocorrélation des composante b_x et b_y du biais $\underline{b}$.

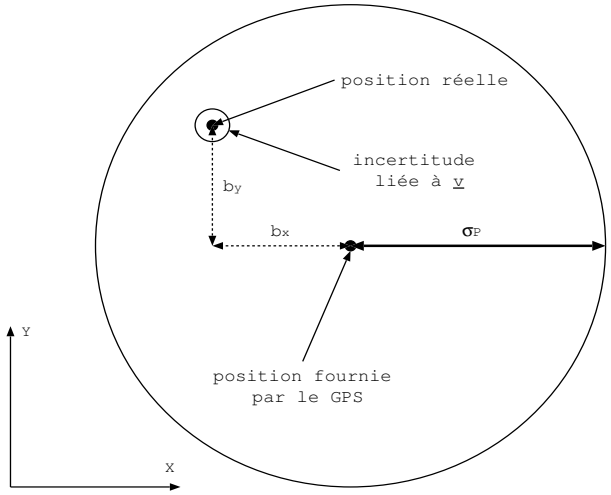


FIG. 5.9 – Modélisation des erreurs GPS.

Le biais est intégré dans le vecteur d'état et l'équation d'observation s'écrit :

$$\underline{X}_{gps} = \mathbf{H}\underline{X}_k + \underline{v} \text{ avec } \mathbf{H} = \begin{pmatrix} 1\,0\,0\,1\,0 \\ 0\,1\,0\,0\,1 \end{pmatrix} \text{ et } \underline{X}_k = \begin{pmatrix} x\ y\ \gamma\ b_x\ b_y \end{pmatrix}^t$$

La matrice de covariance associée à $\underline{X}_{gps}$ est la covariance du vecteur $\underline{v}$:

$$\mathbf{Q}_{gps} = \begin{pmatrix} \sigma^2_{xgps} & 0 \\ 0 & \sigma^2_{ygps} \end{pmatrix}$$

avec σ_{xgps} et σ_{xgps} de l'ordre de $30cm$.

L'estimation du biais n'est pas possible avec la seule observation GPS puisque l'information "biais" n'est pas délivrée par le système GPS. L'hypothèse utilisée (vérifiée par l'expérience) est que le biais reste constant dans un intervalle de temps de $4s$ à $10s$. La détermination du biais nécessite l'utilisation d'une autre information de positionnement absolu. Elle est fournie par le processus de couplage vision carte explicité dans la suite. Ainsi il est possible d'estimer le biais grâce à la fusion de ces deux types d'information avec les données des capteurs proprioceptifs.

Lors de la première observation du GPS les valeurs du biais sont supposées nulles et la sous-matrice $\mathbf{Q}_b$ (matrice de covariance du biais) est initialisée avec la matrice de covariance $\mathbf{Q}i_b$.

$$\mathbf{Q}i_b = \begin{pmatrix} \sigma^2_P & 0 \\ 0 & \sigma^2_P \end{pmatrix}$$

Gestion du biais lors de la mise à jour de l'état

Un filtre de Kalman permet de mettre à jour la position du véhicule ainsi que le biais. Lors de grandes variations du biais il faut vérifier que la mise à jour est possible. La nouvelle valeur du GPS doit être contenue dans l'intervalle de confiance (figure 5.10) décrit par la matrice de covariance $\mathbf{Q}_b$. Un test sur la distance de Mahalanobis entre la position estimée et la position

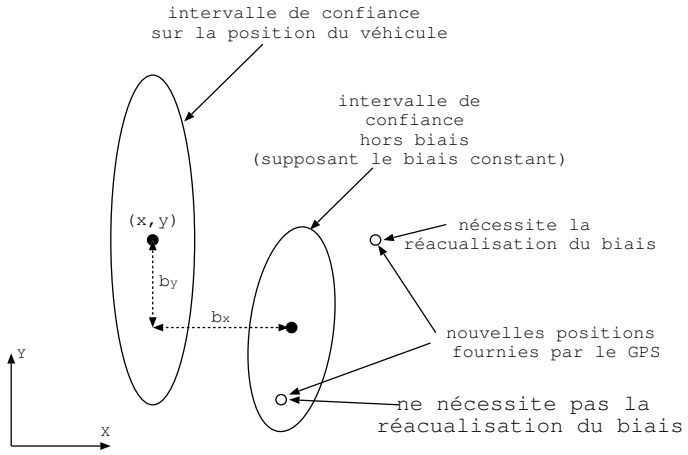


FIG. 5.10 – Test de réactualisation du biais.

biaisée (équation 5.8) permet cette vérification.

$$d = (\underline{X}_{gps} - \mathbf{H}\underline{X}_k)\mathbf{Q}_b^{-1}(\underline{X}_{gps} - \mathbf{H}\underline{X}_k)^t \tag{5.8}$$

Si $d < 1$ le biais est mis à jour sinon il faut réinitialiser la matrice de covariance associée au vecteur d'état avant la mise à jour. Pour ceci $\mathbf{Q}_b$ prend sa valeur initiale ($\mathbf{Q}i_b$) et les covariances entre l'attitude du véhicule et le biais sont annulées.

$\mathbf{Q}$ devient alors :

$$\mathbf{Q} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} & \sigma_{x\gamma} & \\ \sigma_{yx} & \sigma_y^2 & \sigma_{y\gamma} & 0 \\ \sigma_{\gamma x} & \sigma_{\gamma y} & \sigma_\gamma^2 & \\ & 0 & & \mathbf{Q}i_b \end{pmatrix}$$

5.3.3 Localisation par couplage vision/carte

Il s'agit ici dans un premier temps d'effectuer la localisation locale du véhicule sur sa voie de circulation en utilisant la méthode de détection de bords de voies présentée précédemment (§5.1), figure 5.11), puis d'exploiter une cartographie numérique vectorielle de l'environnement routier, qui doit être disponible *a priori*, pour positionner le repère local lié à la route dans le repère global de la carte.

FIG. 5.11 – Détection des bords de voie.

A partir de primitives images représentant le bord de la chaussée, cette méthode de vision permet d'estimer un ensemble de paramètres géométriques de la route ainsi que la position et l'orientation du véhicule sur celle-ci. Une matrice de covariance associée à ces paramètres est aussi estimée. Ces informations sont calculées dans un référentiel tangentiel à la route (voir figure 5.12). Les paramètres estimés sont les suivants : la largeur de la voie de circulation L, la courbure latérale de la route C, l'angle de tangage de la caméra α, la position latérale du véhicule sur la chaussée x_0, l'angle de cap du véhicule ψ.

Modélisation des informations vision Pour la localisation du véhicule sur la chaussée, seuls les deux derniers paramètres (x_0 et ψ)sont utilisables. L'information de position longitudinale est manquante. Elle est impossible à obtenir avec cette approche. Elle est considérée comme nulle avec une confiance très faible. La figure 5.12 montre la représentation de chacun de ces paramètres avec leur intervalle de confiance.

Ces informations vision sont regroupées dans le vecteur $\underline{X}_{vision}$ et dans sa matrice de covariance $\mathbf{Q}_{vision}$ avec :

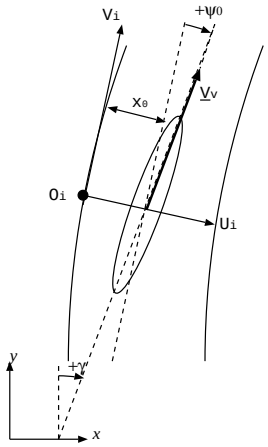


FIG. 5.12 – Représentation des informations vision.

$$\underline{X}_{vision} = \begin{pmatrix} x_0 \\ 0 \\ \psi \end{pmatrix} \text{ et } \mathbf{Q}_{vision} = \begin{pmatrix} \sigma^2_{x_0} & 0 & \sigma_{x_0\psi} \\ 0 & \infty & 0 \\ \sigma_{\psi x_0} & 0 & \sigma^2_{\psi} \end{pmatrix}$$

Couplage vision/carte Les informations données par l'algorithme de vision sont définies dans un repère local lié à la route. Le couplage vision/carte permet de les transposer dans le repère de la carte.

L'origine $\underline{O}$ du référentiel route dans le référentiel monde et la matrice de rotation $\mathbf{M}$ décrivant l'orientation du référentiel route par rapport au référentiel monde, sont nécessaires (figure 5.12) :

$$\underline{O} = \begin{pmatrix} x_{vision} \\ y_{vision} \\ 0 \end{pmatrix} \text{ et } \mathbf{M} = \begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{pmatrix}^t$$

Le vecteur résultant est noté $\underline{X}_{vision/m}$. L'équation de passage du vecteur $\underline{X}_{vision}$ dans le référentiel du monde s'écrit :

$$\underline{X}_{vision/m} = \begin{pmatrix} M_{11}x_0 + x_{vision} \\ M_{21}x_0 + y_{vision} \\ \psi - atan(\frac{M_{12}}{M_{22}}) \end{pmatrix}$$

La matrice de covariance associée au vecteur $\underline{X}_{vision/m}$ est notée $\mathbf{Q}_{vision/m}$.

Modèle d'observation Le modèle d'observation des données vision est décrit par l'équation :

$$\underline{X}_{vision/m} = \mathbf{H_{vis}}\underline{X}_k + \underline{v_{vis}} \text{ avec } \mathbf{H_{vis}} = \begin{pmatrix} 1\,0\,0\,0\,0 \\ 0\,1\,0\,0\,0 \end{pmatrix}$$

L'erreur d'observation est modélisée par un bruit blanc gaussien de covariance $\mathbf{Q}_{vision/m}$. Un nouveau filtre de Kalman permet la mise à jour de l'état.

Localisation relative

Le modèle d'évolution relie par récurrence le vecteur d'état $\underline{X}_{k+1}$ à l'instant $k+1$ aux informations fournies par les capteurs proprioceptifs à l'instant k. Les informations sont l'angle d'orientation des roues δ et le déplacement Δ_d parcouru par le véhicule entre k et $k+1$.

$$\underline{X}_{k+1/k} = f(\underline{X}_{k/k}, \underline{U}_k) \text{ avec } \underline{U}_k = (\Delta_d, \delta).$$

Le modèle utilisé est le modèle dit "bicyclette" (figure 5.13) :

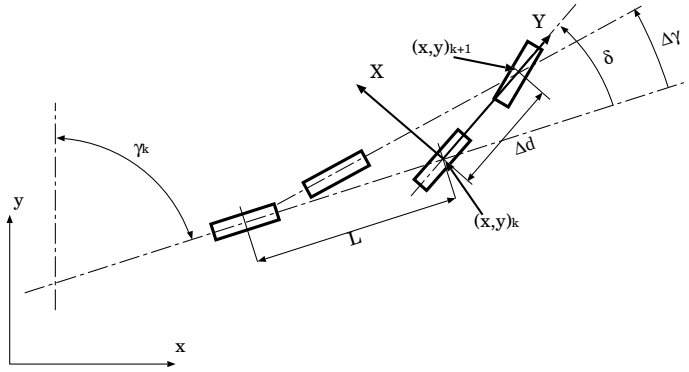


FIG. 5.13 – Modéle "bicyclette".

La position $(x,y)_{k+1}$ est exprimée en fonction de $(x,y)_k$, le véhicule se déplaçant de Δ_d dans l'axe des roues celles-ci étant tournées d'un angle δ.

D'où :

$$\begin{cases} x_{k+1} = x_k - \Delta_d \sin(\gamma_k + \delta) \\ y_{k+1} = y_k + \Delta_d \cos(\gamma_k + \delta) \end{cases}$$

L' orientation du véhicule γ_{k+1} est donnée par la relation :

$$\gamma_{k+1} = \gamma_k + \Delta_{\gamma_k} \text{ avec } \Delta_{\gamma_k} = \arctan \frac{\Delta_d \sin(\delta)}{L + \Delta_d \cos(\delta)}$$

Aucune évolution sur le biais n'est calculable, il est considéré comme fixe d'un instant à l'autre, sauf lors de grandes variations (cf §5.3.2).

Les équations d'évolutions s'écrivent alors :

$$\begin{cases} x_{k+1} = x_k & - \Delta_d \sin(\gamma_k + \delta) \\ y_{k+1} = y_k & + \Delta_d \cos(\gamma_k + \delta) \\ \gamma_{k+1} = \gamma_k & + \Delta_{\gamma_k} \\ b_{x_{k+1}} = b_{x_k} & \\ b_{y_{k+1}} = b_{y_k} & \end{cases}$$

La valeur de γ est estimée modulo 2π. Ceci permet de rester dans un espace borné lors de rotation de plus de 2π (ex : rond point) et d'assurer une compatibilité avec les observations des capteurs extéroceptifs.

La matrice de covariance associée à l'état est actualisée :

$$\mathbf{Q}_{k+1} = \mathbf{J}_e \mathbf{C}_e \mathbf{J}_e^T$$

La matrice $\mathbf{C}_e$ est constituée de $\mathbf{Q}_k$ et de la matrice de covariance $\mathbf{Q}_{U_k}$ du vecteur d'entrée $\underline{U}_k$. $\mathbf{J_e}$ est le jacobien déduit des équations d'évolution.

$$\mathbf{C}_e = \begin{pmatrix} \mathbf{Q}_k & 0 \\ 0 & \mathbf{Q}_{U_k} \end{pmatrix} \text{ avec } \mathbf{Q}_{U_k} = \begin{pmatrix} \sigma^2_{\Delta_d} & 0 \\ 0 & \sigma^2_{\delta} \end{pmatrix} \text{ et } \mathbf{J}_e = \frac{\partial \underline{X}_k}{\partial(x_k, y_k, \gamma_k, \Delta_d, \delta)}$$

La figure 5.14 montre le déplacement du véhicule entre l'instant k et $k+1$ ainsi que les intervalles de confiance sur sa position.

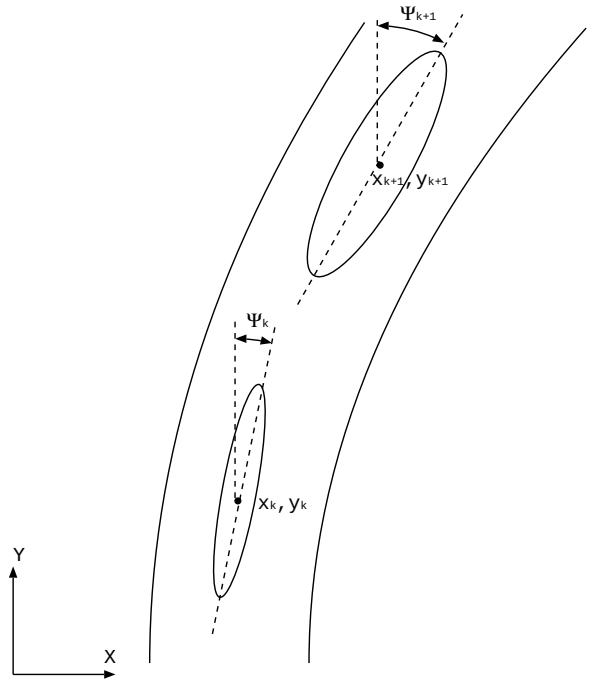


FIG. 5.14 – Evolution du véhicule entre les instants et $k+1$.

5.3.4 Résultat de localisation

Les résultats présentés correspondent à des expérimentations réalisées sur un aménagement 2×2 voies de l'agglomération de Clermont-Ferrand.

La réaction du système de localisation lors de perte de signal GPS est tout d'abord illustrée. Cette situation est assez fréquente en milieu urbain. Afin de démontrer la robustesse de l'approche, une perte de signal GPS de $30s$ a été simulée dans deux contextes différents : avec ou sans localisation par vision (voir figure 5.15).

La figure 5.15(a) est le résultat de la localisation sans vision et avec peu de perte de signal GPS. Elle montre une trajectoire du véhicule qui passe par l'ensemble des données GPS car sans vision, les informations GPS sont les plus précises et l'estimation du biais est impossible. La figure 5.15(b) est le résultat de la localisation sans vision et avec une longue perte de signal GPS. Plus l'absence de signal GPS est longue, plus la dérive de la position par rapport au résultat du premier graphique est importante (comme énoncé §5.3.2). La figure 5.15(c) est le résultat de la localisation avec vision. Pendant la perte de signal GPS, le système continue de fournir une position correcte du véhicule. La trajectoire du véhicule ne passe pas par l'ensemble des points GPS : le système est capable d'estimer le biais.

Aucune localisation "réputée vraie" n'est disponible pour servir de référence afin de comparer rigoureusement celle qui est estimée par le système proposé. L'erreur d'estimation n'est pas possible à obtenir. L'évaluation de la précision est faite en analysant l'évolution des écart-types *a posteriori* sur les paramètres de localisation estimés dans le référentiel de la route(figure 5.16).

De $t = 0$ à $t = t_0$, la localisation par vision est inhibée, seules la localisation par GPS et la localisation relative fonctionnent. Les erreurs sur la position latérale et la position longitudinale sont de l'ordre de $25m$ ($\simeq$ précision du GPS). L'erreur sur γ est de l'ordre de 4 degré avec des fluctuations dues aux incertitudes sur les capteurs proprioceptifs.

De t_0 à t_1 la vision est validée. L'ensemble des erreurs diminuent, l'erreur latérale décroît jusqu'à $20cm$,l'erreur longitudinale jusqu'à $4m$ et l'erreur sur γ jusqu'à environ 2 degrés. Cette valeur moyenne de l'erreur peut être interprétée comme un biais sur l'orientation de la caméra.

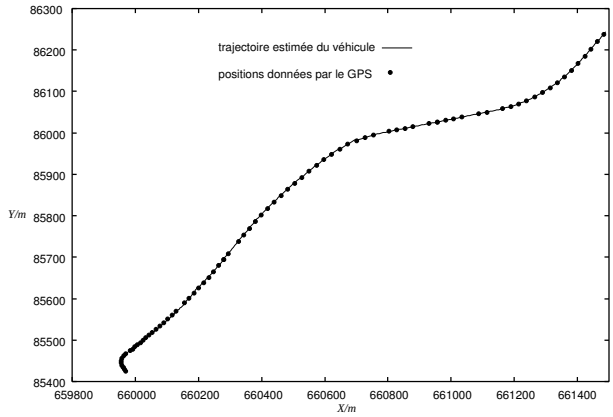
De t_1 à t_2 la perte de signal GPS n'affecte pas les résultats sur la localisation. Des variations assez fortes sur γ sont visibles, elles sont dues aux changements de voie de circulation ou à la non reconnaissance de la route par l'algorithme de vision.

Après t_2 il n'y a pas de changement significatif sur les résultats.

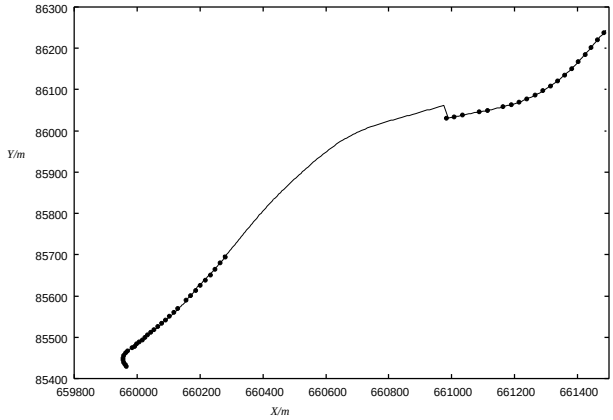
Conclusion

Une méthode de fusion de données multicapteurs (proprioceptifs et extéroceptifs) permettant de localiser un véhicule dans un environnement cartographié est présentée dans cette partie. Elle utilise un GPS bas coût avec une précision de $30m$. Les résultats montrent la localisation latérale du véhicule sur la chaussée avec une précision minimale de $20cm$. Elle met en œuvre le système de perception proposé pour réaliser l'analyse des images en vue d'une localisation par vision. La modélisation complète des erreurs GPS est également réalisée. Cette approche originale a été valorisée par 2 articles et 5 communications internationales.

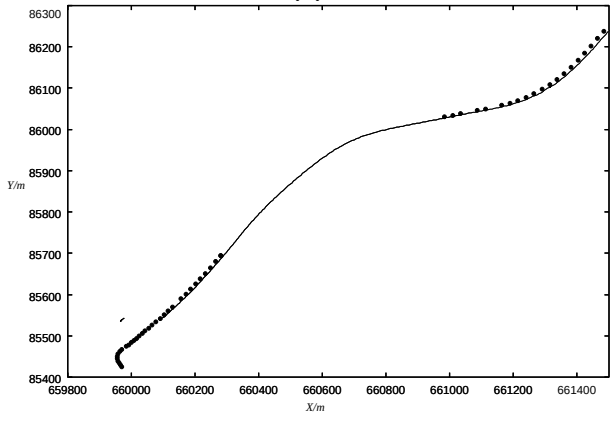
Le système de perception attentionnelle apparaît dans cette approche de manière sous-jacente dans la localisation par couplage vision/carte. Le processus attentionnel s'appuie sur un modèle générique de route pour focaliser la détection des bords de voie dans l'image. Il ne subit aucune adaptation majeure dans cette approche. L'utilisation de la carte pour transposer la localisation locale dans un repère absolu s'effectue ensuite. Il apparaît donc que la carte ne joue pas un rôle explicite dans le processus de perception. Or celui-ci à besoin d'un modèle de l'environnement pour guider la focalisation des détections dans les données provenant des capteurs extéroceptifs. L'hypothèse que la carte numérique puisse jouer ce rôle de modèle de l'environnement ne parait pas aberrante. C'est sur elle que repose l'application présentée dans la section suivante.



(a)



(b)



(c)

FIG. 5.15 – Influence d'une perte de signal GPS : (a) localisation sans interruption des informations GPS, (b) la localisation sans la vision avec une longue interruption du GPS fait apparaître une dérive importante du positionnement, (c) avec la fusion GPS/vision/capteurs proprioceptifs la localisation ne dérive pas.

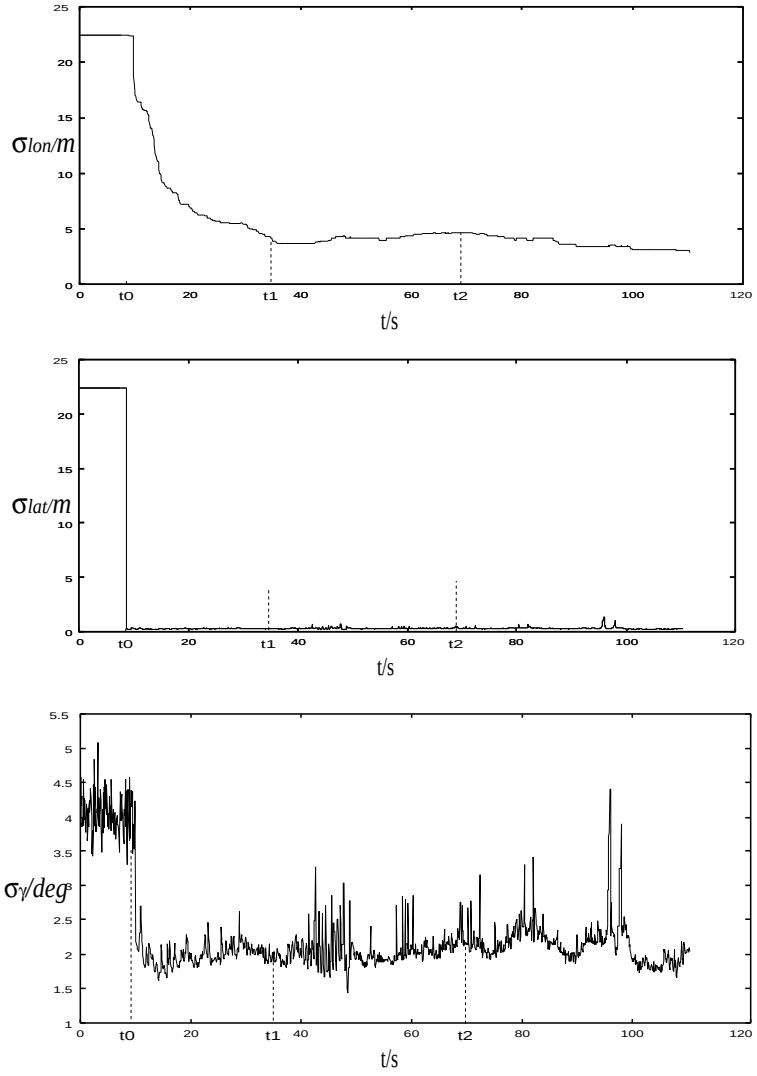


FIG. 5.16 – Erreur sur les paramètres de localisation.

5.4 Localisation et guidage d'un robot mobile autonome

5.4.1 Introduction

La localisation peut être définie comme le processus qui consiste à déterminer la position du robot par rapport à des points de repère connus qui constituent des amers. La méthode décrite à la section5.3 met en oeuvre de manière juxtaposée un système de perception attentionnel guidé par un modèle générique des amers (les bords de voies) exprimés dans un repère local et une carte précise de ces mêmes amers dans un repère absolu. L'idée de guider le processus de perception attentionnel non plus par le modèle générique mais directement par la carte numérique des amers apparaît comme une évolution naturelle des travaux précédents. De plus l'utilisation d'un seul type de capteur extéroceptif limite la nature des amers que le processus de localisation est susceptible de percevoir. C'est pourquoi, le système de perception attentionnelle des amers a été généralisé à un contexte multisensoriel. Cette application est l'objet des travaux de thèse de Cédric Tessier soutenue le 9 novembre 2007 [21].

La carte des amers constitue la représentation synthétique de la scène qui guide leur observation. Elle comprend la position géométrique des amers. Cette carte est complétée d'autres types d'information : chaque amer est associé aux capteurs extéroceptifs qui peuvent permettre de l'observer et aux algorithmes (détecteurs) capables d'extraire une observation de l'amer dans les données capteur. L'association d'un amer, d'un capteur et d'un détecteur constitue un "triplet perceptif". L'introduction de cette notion nouvelle est une contribution importante de ce travail. La représentation synthétique qui guide la perception est donc une carte de triplets perceptifs. Le processus attentionnel peut alors sélectionner à chaque instant le meilleur triplet à observer c'est-à-dire celui qui contribue le mieux à la compréhension de la scène (ici essentiellement à la localisation du robot). Même suffisamment focalisée, la recherche peut encore associer une observation avec un amer qui ne lui correspond pas : les erreurs d'appariement même limitées existent toujours. Le processus intègre le calcul d'un critère de confiance sur l'estimée courante de la pose du véhicule. La vraisemblance de l'hypothèse courante de localisation est mise à jour après chaque détection. En cas de faux appariement, le processus récursif de focalisation définit des zones d'intérêt où la détection attendue des amers échoue. La probabilité de l'hypothèse courante de localisation chute. Le processus de remise à jour remonte alors la chaîne des hypothèses jusqu'à la dernière valide et en initie une nouvelle avec une autre hypothèse d'appariement amer/observation. Ce processus apporte une contribution importante et originale au calcul de l'intégrité des estimations de localisation, problème qui, s'il est bien identifié par la communauté semble le plus souvent éludé.

5.4.2 Intérêt de l'application

L'augmentation de l'autonomie des robots mobiles, avec en ligne de mire la réduction de l'intervention d'un opérateur humain, est un problème souvent abordé tant par la recherche scientifique que par les industriels. De tels systèmes sont envisagés par exemple pour des trajets touristiques, pour des véhicules de nettoyage ou encore pour le guidage d'engins agricoles,. . . Certains sont déjà opérationnels comme par exemple dans le domaine agricole, les systèmes proposés par CLAAS [15], John Deere [64], [70]. Cependant ils ont été conçus pour réaliser des tâches spécifiques comme la moisson [15], et réussissent parfaitement des trajets rectilignes [64].

L'augmentation de leur autonomie passe nécessairement par l'élargissement de leur capa-

cité de déplacement automatique. Ils doivent ainsi être capables de suivre des trajectoires quelconques avec une bonne précision et le système de localisation embarqué doit donc être performant en termes de précision comme d'intégrité.

Cette partie aborde le problème de la localisation d'un robot mobile (ou d'un véhicule) en vue de son guidage automatique. La solution proposée repose sur l'utilisation d'une carte de l'environnement : c'est une approche de type MAL (Map Aided Localization). La localisation repose sur la perception des amers contenus dans la carte par des capteurs extéroceptifs embarquées. La carte est connue a priori.

Dans ce contexte, les capteurs utilisés sont soit des télémètres LASER (LIDAR) [10], [59], [32], [18], soit des caméras [50], [28], [24], soit les deux [27]. Dans ce dernier cas, la solution est multisensorielle, ce qui élargit le champ de vue du robot depuis une position donnée. Ceci signifie non seulement que la perception couvre une zone géographique plus vaste mais aussi que des amers de différents types peuvent être observés le long de la trajectoire à suivre. Le système de localisation peut être utilisé dans des cas plus divers et la précision de localisation est maintenue de manière plus constante. La solution décrite ici est elle-même multisensorielle et associe un LIDAR et une caméra. Des capteurs proprioceptifs sont également utilisés.

Dans ce contexte le principal problème concerne l'association de données [62], [60] : il s'agit de mettre en correspondance un amer (information contenue dans la carte) et une primitive extraite des données issues des capteurs (une observation). Un tel appariement amer/observation est correct si l'observation est associée à l'amer qui lui correspond réellement. Deux types d'appariement conduisent à une mauvaise association de données : un amer est associé avec une observation qui ne lui correspond pas ou un amer est associé avec du bruit.

La recherche active des amers est une solution originale introduite par Fox [33] pour résoudre ce problème et utilisée ensuite par Davison [24]. Elle consiste à ne pas rechercher les observations dans la totalité des données capteurs mais dans des zones d'intérêt. Celles-ci sont définies idéalement de telle sorte que l'observation correspondant à un amer se trouve inévitablement dans la zone lorsque la position estimée du robot est correcte. Elles sont construites à partir de la valeur estimée courante de la pose (position et attitude) du robot et de son incertitude, ainsi qu'à partir de la position de l'amer contenue dans la carte (Fig. 5.17).

Ainsi, comme l'algorithme de détection extrait l'observation liée à un amer dans une zone d'intérêt, le rapport signal sur bruit augmente et la probabilité de mauvais appariement chute.

Finalement, l'observation extraite des données capteurs par un algorithme de détection adéquat, est associée à l'amer recherché et la pose du véhicule est mise à jour.

Dans l'approche par recherche active, la combinatoire du problème d'appariement est *a priori* de 1 vers n ce qui signifie qu'un amer peut être associé aux n observations retournées par l'algorithme d'extraction appliqué dans la zone d'intérêt. Si, comme dans la solution proposée dans cette partie, l'algorithme d'extraction retourne une seule observation, c'est nécessairement elle qui est appariée avec l'amer. Le risque de faux appariement est plus élevé que si plusieurs observations sont extraites dans une même zone d'intérêt. Dans ce dernier cas en effet, la probabilité que l'observation correcte soit dans la liste est plus importante que dans le premier où il n'y en a qu'une. La contrepartie est un temps de calcul plus long dans le second cas puisque qu'il nécéssite le traitement de plusieurs hypothèses d'appariement.

Or comme dans les méthodes par recherche active d'amers, la zone d'intérêt est déterminée à partir de la pose courante, un seul mauvais appariement fait diverger la mise à jour de la pose et finalement le processus échoue. Ce problème n'est pas ou peu abordé par la communauté

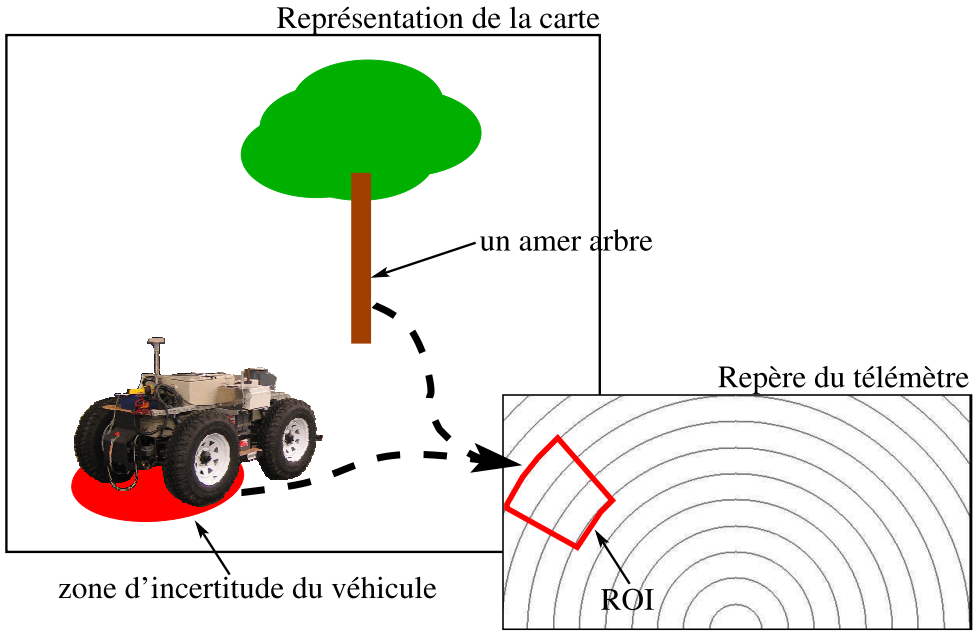


FIG. 5.17 – Illustration du principe de la recherche active des amers : une zone d'intérêt (ROI) correspondant à un amer de type arbre est définie dans l'espace des données du capteur LIDAR.

dans le cas de la localisation à partir d'une carte. Il est en revanche largement traité dans les techniques de cartographie automatique [61], [53] et [86] pour lesquelles plusieurs solutions sont proposées. Pour la localisation/cartographie simultanée (SLAM[2]) un processus efficace d'appariement doit établir la différence entre le bruit et les observations nouvelles, en plus de la fonction de base d'association entre les primitives déjà disponibles avec les observations. La méthode la plus utilisée dans le cas du SLAM est l'algorithme du plus proche voisin [34]. Il associe une primitive à l'observation la plus proche dans une région de validité sur la base d'une mesure de distance (classiquement la distance de Mahalanobis). La facilité de mise en œuvre de cette solution est contre-balancée par des performances limitées principalement en raison du fait que les décisions d'appariement une fois prises ne peuvent être annulées. Dans l'approche Graph-SLAM [79], les appariements sont contrôlés par un arbre d'association de données. Toutes les hypothèses d'appariement possibles sont suivies et finalement seule la plus probable est retenue.

La méthode du suivi multi-hypothèses (MHT) [47], appliquée au cas de l'appariement amer/observation consiste à suspendre la décision d'apparier jusqu'à ce qu'un nombre suffisant d'observations supplémentaires soient faites. Ce processus permet de corriger *a posteriori* les faux appariements. La décision est différée en cas de situation conflictuelle. Toutes les hypothèses possibles d'appariement forment un arbre dont les branches sont suivies en même temps. Les hypothèses sont propagées vers les itérations suivantes avec l'espoir qu'une nouvelle observation résoudra le conflit. Cette méthode permet donc de gérer les observations manquées ou fausses (bruit). Le principal inconvénient est que l'arbre des hypothèses croît exponentiellement au cours du temps ce qui rend l'approche inutilisable à la cadence nécessaire pour un guidage automatique. Le FastSLAM [62] est une implémentation de cette méthode utilisant un filtre à particules.

[2]Simultaneous Localization And Mapping

5.4.3 Principe de la méthode de localisation utilisant la focalisation progressive

Géneralités

Le système de guidage d'un robot mobile autonome proposé estime périodiquement la pose du robot. A chaque estimation, la direction et la vitesse du robot sont ajustées pour reproduire une trajectoire préalablement enregistrée ou définie hors ligne (Fig. 5.18).

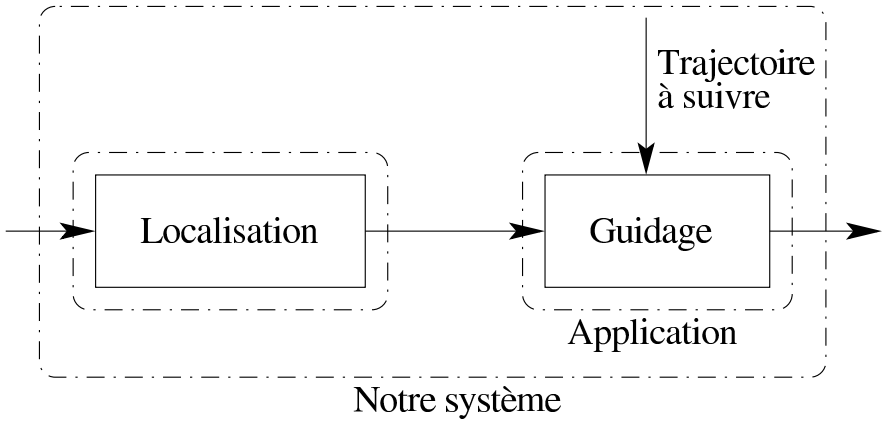


FIG. 5.18 – Principe de la méthode proposée.

La solution proposée est multisensorielle. Elle associe les capteurs embarqués suivants :
– un récepteur GPS bas coût,
– des capteurs extéroceptifs : une caméra vidéo et un télémètre LASER pour la perception des amers,
– des capteurs proprioceptifs : un odomètre et un gyromètre pour une localisation relative.

Le système proposé est de type MAL. La carte est un Système d'Information Géographique (SIG) qui répertorie les amers de l'environnement. Chaque amer (arbre, mur, bord de voie, ...) est placé dans la carte. Il est caractérisé par des paramètres : nom, type, position ... A partir de la pose du robot, le système choisit dans la carte un amer à percevoir et définit une zone d'intérêt dans les données capteur où cet amer est susceptible d'être observé. Un détecteur est alors appliqué à cette zone d'intérêt pour en extraire l'observation. L'appariement de l'amer et de l'observation détectée est alors réalisé de fait (Fig. 5.19).

Triplets perceptifs

Le choix d'une solution multisensorielle permet d'augmenter le nombre et la variété des amers visibles à un instant donné. Dans certains cas un même amer peut être perçu par des capteurs différents. Pour un même capteur, l'observation de l'amer peut être extraite par plusieurs algorithmes de détection.

A priori, les couples capteur/algorithme de détection n'ont pas la même performance de reconnaissance. Par exemple certains sont plus précis et d'autres plus fiables, certains calculés plus rapidement et d'autres plus lentement. De plus, en raison de la position géométrique de l'amer, tous n'améliorent pas l'estimation de la localisation du robot de la même manière. Par exemple, à un moment donné, un amer est utile pour estimer la direction du robot et à un autre moment, il est utilisé pour corriger la position. Le système qui est proposé, prend explicitement

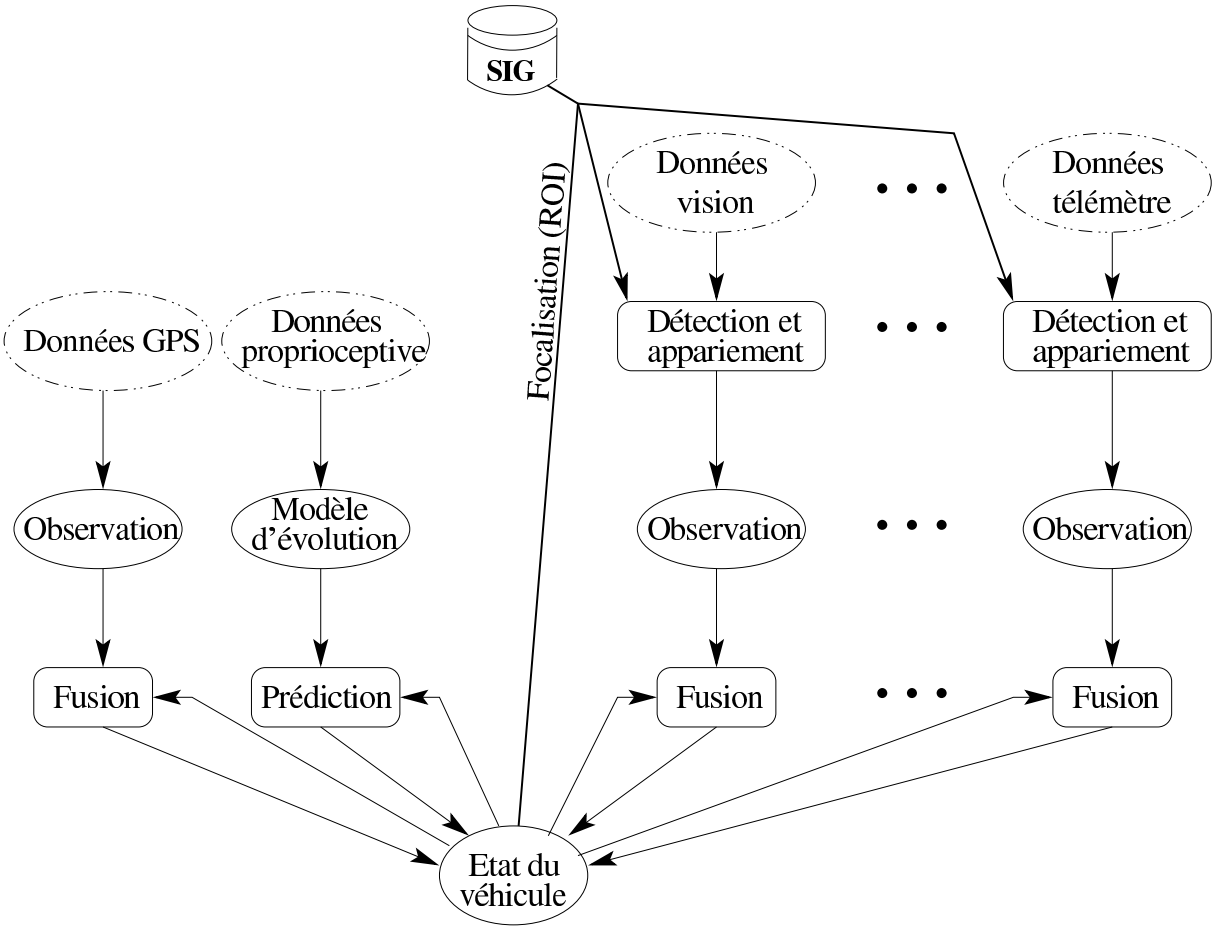


FIG. 5.19 – Schéma du principe de localisation.

en compte ce type de critère pour la perception des amers grâce à la notion de **triplets perceptifs**. Un triplet perceptif est une entité autonome en charge de l'observation d'un amer donné dans une zone d'intérêt définie par la méthode de recherche active, utilisant un capteur donné et un algorithme de détection donné. Le but du système n'est alors plus tant de sélectionner un amer à observer que de choisir le triplet perceptif le plus à même d'améliorer la précision de la localisation.

L'analyse d'un scénario de localisation permet d'illustrer l'utilisation des triplets perceptifs. La figure 5.20 montre la carte vue de dessus. Elle contient cinq amers : trois arbres et deux bords de voie.

Dans le but de localiser le robot, sa position est initialisée avec un GPS. Cette première estimation de l'état est appelée *Etat 0* sur la figure 5.21.

A partir de la pose du robot, le système de localisation construit des triplets perceptifs. Chaque triplet a en charge la détection d'un amer particulier dans une zone d'intérêt à l'aide d'un capteur particulier et de l'algorithme de détection associé. Cet ensemble de triplets est représenté sur la deuxième ligne de la figure 5.21. Le système sélectionne alors le triplet à utiliser. Supposons que le triplet 1 soit sélectionné. L'algorithme de détection du triplet 1 est appliqué dans la fenêtre d'intérêt du capteur associé au triplet 1. Il retourne une observation qui est prise en compte par le système pour remettre à jour la position du robot qui se trouve alors dans l'*Etat 1*. Alors, pour chaque triplet restant, la zone d'intérêt de l'amer associé est ré-évaluée à partir

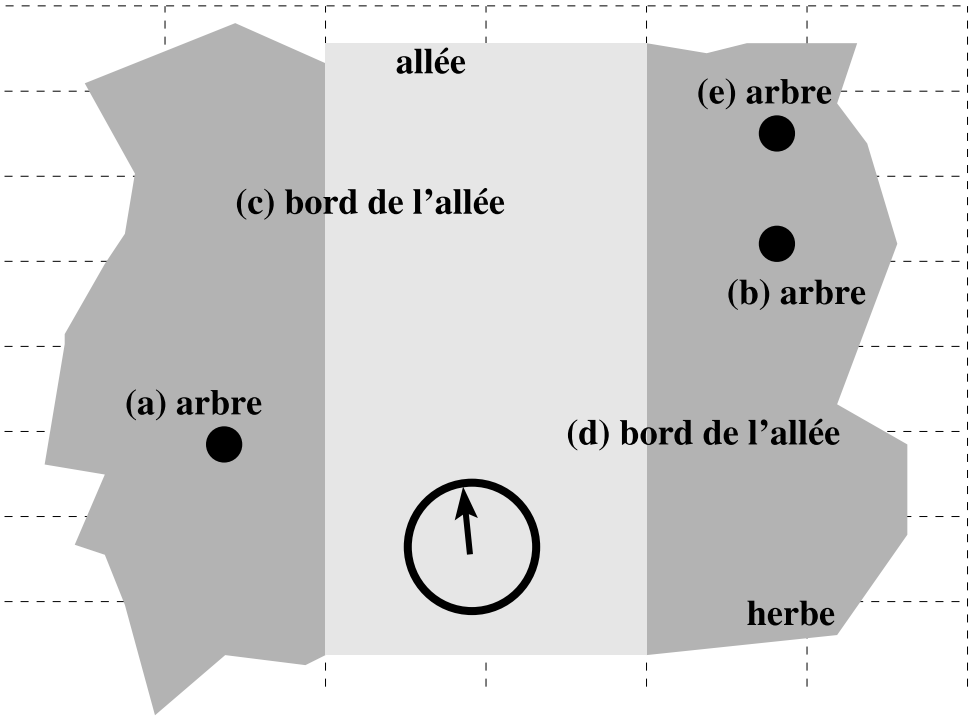


FIG. 5.20 – Vue de dessus de la carte avec trois arbres et deux bords de voie. Le robot est représenté par un cercle.

de ce nouvel état. Ce processus est réitéré jusqu'à ce que la précision requise par l'application soit atteinte ou que tous les triplets aient été traités.

5.4.4 Traitement probabiliste des mauvaises associations de données

Chaque triplet a en charge la détection de l'observation liée à un amer dans une zone d'intérêt. L'association amer/observation est donc implicite. Elle est à la base de la remise à jour de la pose. Si pour une raison ou pour une autre cette association de données n'est pas correcte, la position estimée ne l'est pas non plus. Concrètement, cela signifie que la position "réelle" du robot ne se situe pas dans la zone d'incertitude de l'estimation définie comme l'ellipse à un écart-type autour de l'estimation. Les triplets restants voient alors leur zone d'intérêt mise à jour à partir d'un état incorrect. Elles sont donc nécessairement mal positionnées et ne contiennent probablement pas l'amer. La détection des autres amers dans leur zone mal placée échoue donc et le système n'est pas capable de localiser le robot.

Le calcul d'un niveau de confiance dans l'estimation courante de l'état du robot est proposé pour résoudre ce problème. La non détection d'un amer, qui témoigne d'une estimation incorrecte due à une mauvaise association de données, entraîne la baisse de ce niveau de confiance. Lorsqu'il devient trop bas, le système prévoit de remonter les étapes de mise à jour de l'estimation jusqu'à la dernière valide, puis de recommencer à partir d'une autre hypothèse de correspondance amer/observation.

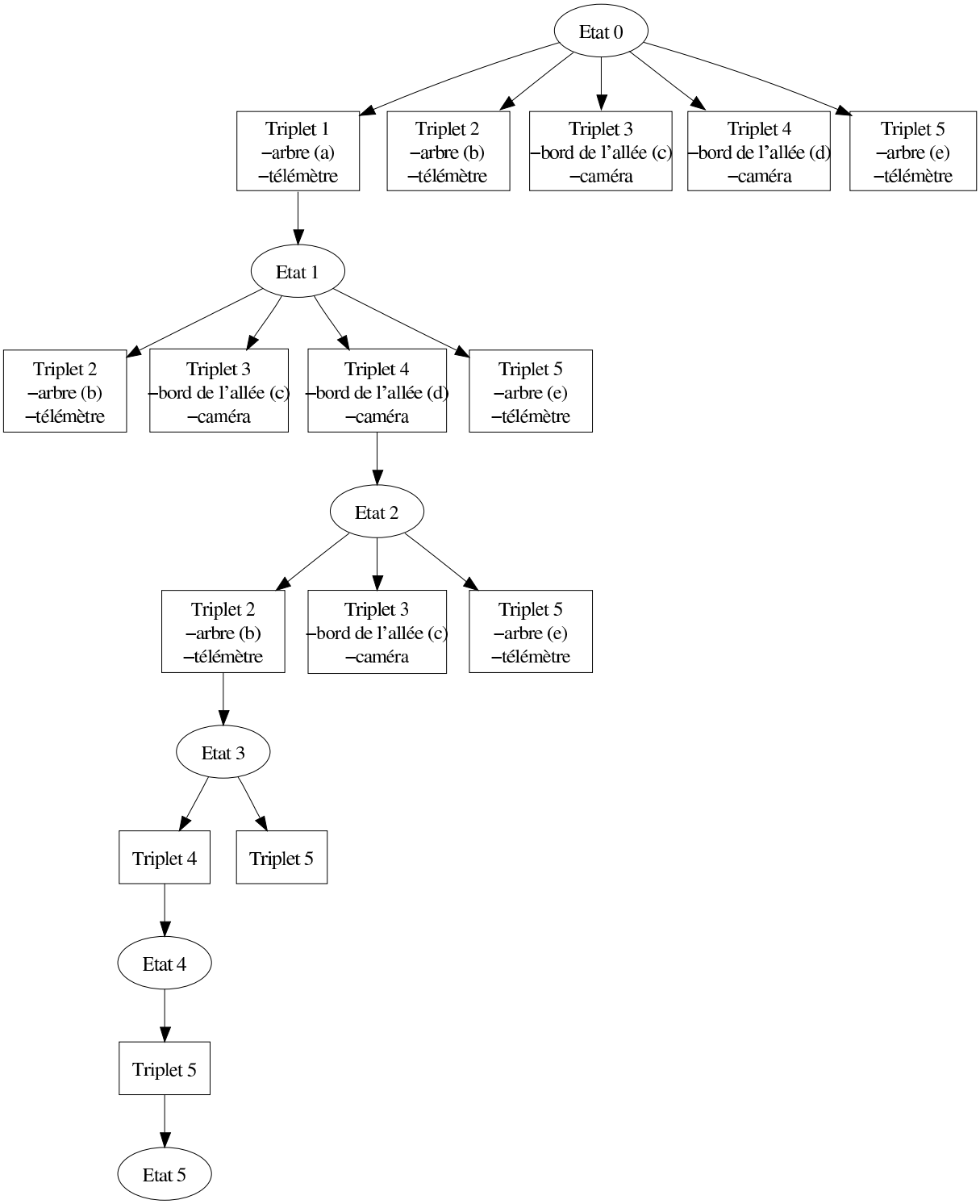


FIG. 5.21 – Evolution de la perception des amers : choix successif des triplets perceptifs dans le cas d'un scénario de localisation.

***Niveau de confiance* associé à l'estimation courante de la pose du robot.**

L'état $\underline{X}$ et sa covariance $\mathbf{C_X}$ sont estimés par un EKF[3] :

$$\underline{X} = (x, y, \theta, v, \omega)^t \tag{5.9}$$

avec :
– x abscisse de la position dans le repère absolu,
– y ordonnée de la position dans le repère absolu,
– θ angle de cap de la position dans le repère absolu,
– v vitesse linéaire,
– ω vitesse angulaire.

La figure 5.22 illustre et définit les notations utilisées.

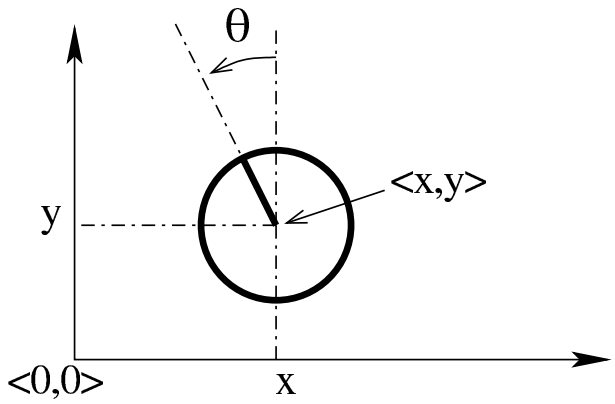


FIG. 5.22 – Pose du robot dans une repère global.

Le *niveau de confiance* est défini par la probabilité de l'événement **H** : *La position réelle du robot est située dans son ellipse d'incertitude à un écart-type autour de l'estimée courante de la position*.

Dans la suite l'expression "l'état est valide" est utilisée pour signifier que l'événement **H** est vrai.

En résumé, le système de localisation proposé estime périodiquement l'état du véhicule :
– le vecteur d'état $\underline{X}$ et sa matrice de covariance $\mathbf{C_X}$ (utilisation d'un EKF et à partir de la recherche active des amers),
– le niveau de confiance (c'est-à-dire la probabilité de l'événement **H**) (formalisation probabiliste détaillée dans les parties suivantes).

Evolution du niveau de confiance

Le système de localisation utilise un filtre bayésien (EKF). Il est donc composé de deux étapes : la prédiction et la mise à jour. Cette partie décrit la manière dont chacune de ces deux étapes influe sur le calcul de la probabilité de l'événement **H**.

[3]Extended Kalman Filter

Etape de prédiction Cette étape consiste à prédire l'état futur du véhicule à partir de l'état courant (Figure 5.23).

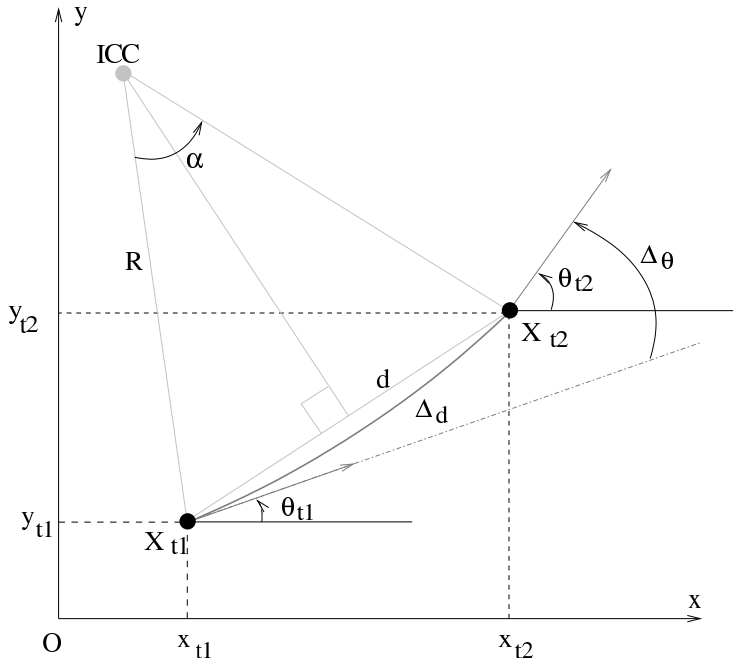


FIG. 5.23 – Petit déplacement entre deux positions successives. La position à la date $t1$ est représentée par X_{t1} et la position à la date $t2$ par X_{t2}.

Pour cela, un modèle d'évolution stochastique est utilisé (Eq.5.10) :

$$\begin{bmatrix} x \\ y \\ \theta \\ v \\ \omega \end{bmatrix}_{t2} = \begin{bmatrix} x \\ y \\ \theta \\ v \\ \omega \end{bmatrix}_{t1} \begin{matrix} -\Delta_d \cdot sin(\theta_{t1} + \Delta_\theta/2) \\ +\Delta_d \cdot cos(\theta_{t1} + \Delta_\theta/2) \\ +\Delta_\theta \qquad\qquad +\underline{b} \end{matrix} \tag{5.10}$$

où :
– $\Delta_t = t2 - t1$,
– $\Delta_d = \Delta_t \cdot v_{t1}$,
– $\Delta_\theta = \Delta_t \cdot \omega_{t1}$,
– b est un bruit blanc gaussien qui est caractéristique de l'erreur de modélisation.
D'après cette hypothèse, le modèle d'évolution représente de façon réaliste le déplacement du véhicule, uniquement dans le cas où l'état est valide à l'instant $t1$, il sera également valide à l'instant $t2$. De plus, si l'état n'est pas valide à la date $t1$ il a une probabilité élevée de ne pas l'être non plus à la date $t2$. En conséquence la probabilité $P(\mathbf{H})$ n'évolue pas pendant l'étape de prédiction :

$$\textbf{Etape de prédiction :} \quad P(\mathbf{H})_{t2} = P(\mathbf{H})_{t1} \tag{5.11}$$

Etape de mise à jour L'étape de mise à jour consiste à actualiser l'état du véhicule à partir d'une observation. Une observation est fournie par un triplet perceptif. La position du véhicule est mise à jour à partir des informations issues des capteurs proprioceptifs ou extéroceptifs.

Or pour un capteur proprioceptif, il n'y a pas de problème d'association de données [19]). Le niveau de confiance n'évolue donc pas :

$$\textbf{Capteur proprioceptif :}\ P(\mathbf{H})_{\text{état mis à jour}} = P(\mathbf{H})_{\text{état prédit}} \tag{5.12}$$

Les capteurs extéroceptifs sont utilisés pour détecter les amers dans l'environnement du véhicule. L'observation d'un amer dans les données capteur peut ne pas correspondre à cet amer. En effet, le système peut avoir détecté un autre amer ou bien du bruit. Ceci occasionne une mauvaise association de donnée. Pour réduire ce problème de mauvais appariement, le système de localisation proposé utilise le processus attentionnel de focalisation progressive

Deux propositions sont définies :

Proposition 1 : si l'amer ne peut être détecté dans sa zone d'intérêt alors l'état courant n'est pas valide.

Proposition 2 : Le contraire de la proposition 1 est faux. En effet, même si l'état n'est pas valide, un amer peut se trouver dans sa zone d'intérêt car la fonction de projection d'un amer du repère du GIS vers le repère des données capteurs n'est pas toujours bijective : plusieurs positions du véhicule peuvent conduire à la même zone d'intérêt pour un amer donné.

L'évolution du niveau de confiance repose sur les propositions précédentes c'est-à-dire sur des notions d'information "positive" ou "négative". Une information positive concerne la détection d'un amer cependant qu'une information négative concerne l'échec de la détection d'un amer. En général les systèmes de localisation ne prennent pas en compte ces informations négatives. Par exemple, le fait de ne pas voir la tour Eiffel à Paris alors qu'elle est activement recherchée, rend extrêmement improbable le fait d'être situé tout prés d'elle !

Ici, la détection d'un amer est réalisée par un triplet perceptif. Quand l'état du véhicule est valide, le triplet peut retourner plusieurs résultats lorsqu'il recherche un amer particulier dans sa zone d'intérêt :

- **la détection de l'amer recherché réussit,** l'information détectée correspond à du bruit ou à un autre amer, le système considère que l'amer correct est détecté (même s'il est toujours possible qu'il corresponde à du bruit c'est-à-dire au résultat d'une erreur de détection),
- **la détection de l'amer échoue** (aucun résultat n'est retourné par le triplet) cependant que l'état du système est valide.

Lorsque l'état du véhicule est valide l'algorithme peut :
- ne retourner aucun résultat,
- détecter une information (bruit).

Le fait qu'un triplet retourne un résultat, ne signifie pas que l'amer recherché est dans la zone d'intérêt puisqu'il peut toujours s'agir de bruit. De la même manière le fait qu'il ne retourne aucun résultat ne signifie pas que l'état est invalide car les détecteurs ne sont jamais parfaits.

C'est pourquoi le niveau de confiance est modifié de manière stochastique :

1. Le succès d'une détection renforce la confiance dans l'estimation courante de la position et le niveau de confiance augmente. La détection est prise en compte pour la mise à jour de l'estimation du vecteur d'état, de sa covariance et du niveau de confiance.

Capteur extéroceptif (Détection réussie) :
$P(\mathbf{H})_{\text{état courant}}$: augmente
$P(\mathbf{H})_{\text{nouvel état}}$: fonction de $P(\mathbf{H})_{\text{état courant}}$

(5.13)

2. L'échec d'une détection affaiblit la confiance dans l'estimation courante de la position et le niveau de confiance diminue.

Capteur extéroceptif (Détection échoue) :
$P(\mathbf{H})_{\text{état courant}}$: diminue

(5.14)

Calcul du niveau de confiance

Notations Les notations suivantes sont utilisées :
– k : le temps discrétisé
– u_k : mesure d'odométrie entre les instants $k-1$ et k
– y_k : résultat d'un triplet
– $\underline{X_k}$: vecteur d'état décrivant la pose vraie du véhicule à l'instant k
– h_k : variable aléatoire correspondant à l'événement $\mathbf{H}$ à l'instant k

Principe Le calcul du niveau de confiance peut être formalisé par un réseau bayésien dynamique (Figure 5.24) avec pour objectif le calcul de la probabilité :

$$P(h_k \mid y_{1:k},\ u_{1:k}) \tag{5.15}$$

Cependant, comme cela est expliqué précédemment le déplacement du véhicule n'affecte pas le niveau de confiance. L'équation (5.15) devient :

$$P(h_k \mid y_{1:k}) \tag{5.16}$$

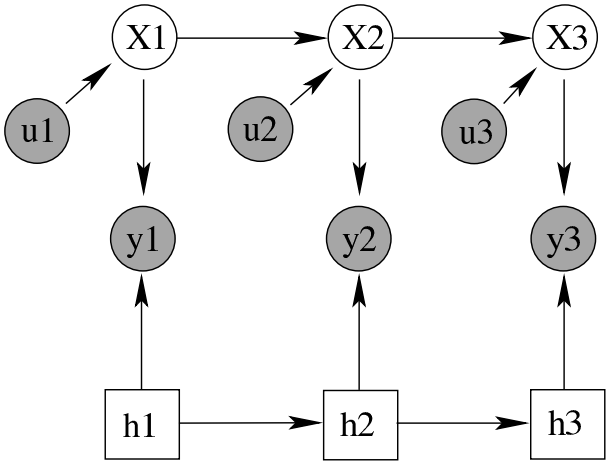


FIG. 5.24 – Réseau bayésien dynamique pour gérer le niveau de confiance h_k.

Les équations 5.13 et 5.14 montrent que la mise à jour du niveau de confiance est fonction de la probabilité que le triplet perceptif retourne un résultat. Pour chaque triplet, deux termes sont définis :
– $P(y_k|h_k)$ $(P(\overline{y}_k|h_k))$: représente la probabilité que le triplet retourne (ou pas) un résultat quand l'état du véhicule est valide.
– $P(y_k|\overline{h}_k)$ $(P(\overline{y}_k|\overline{h}_k))$: représente la probabilité que le triplet retourne (ou pas) un résultat quand l'état du véhicule n'est pas valide.

Ces deux probabilités sont calculées en tenant compte du capteur, du détecteur, de l'amer et surtout de la taille de sa zone d'intérêt. Lorsque la focalisation est resserrée (zone d'intérêt de taille réduite) $P(y_k|h_k)$ est très supérieure à $P(y_k|\overline{h}_k)$. Grâce à l'inférence bayésienne, le niveau de confiance est calculé de la manière suivante :

– quand le triplet réussit à détecter un amer (pas forcément celui qui est recherché)

$$P(h_k|y_k) = \frac{P(y_k|h_k) \cdot P(h_k)}{P(y_k)} \tag{5.17}$$

avec : $P(y_k) = P(y_k|h_k) \cdot P(h_k) + P(y_k|\overline{h}_k) \cdot P(\overline{h}_k)$

– quand le triplet ne réussit pas à détecter un amer (i.e. : aucun résultat retourné) :

$$P(h_k|\overline{y}_k) = \frac{P(\overline{y}_k|h_k) \cdot P(h_k)}{P(\overline{y}_k)} \tag{5.18}$$

avec : $P(\overline{y}_k) = P(\overline{y}_k|h_k) \cdot P(h_k) + P(\overline{y}_k|\overline{h}_k) \cdot P(\overline{h}_k)$

Quand la détection réussit (l'événement y_k est réalisé), la position du véhicule est mise à jour (figure 5.25). Le niveau de confiance $P(h_{k+1})$ du nouvel état est également mis à jour en prenant en compte la probabilité que l'association de données soit correcte.

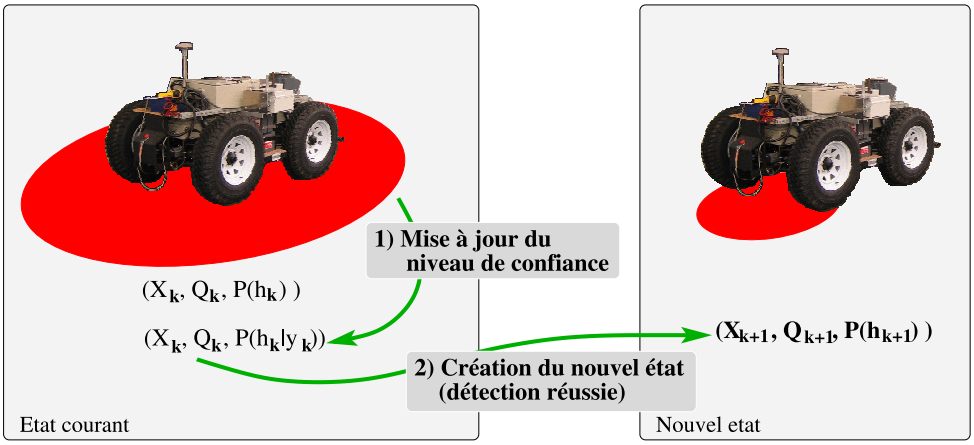


FIG. 5.25 – Présentation des deux étapes qui suivent une détection réussie. La première est la mise à jour du niveau de confiance et la seconde celle de l'état.

En fait, cette nouvelle probabilité correspond à la probabilité conjointe de toutes les hypothèses de mise en correspondance :

$$\begin{aligned} P(h_{k+1}) &= P(h_{k+1}, h_k, ..., h_1) \\ &= P(h_{k+1}|h_k) \cdot P(h_k) + P(h_{k+1}|\overline{h}_k) \cdot P(\overline{h}_k) \end{aligned} \tag{5.19}$$

où

– $P(h_{k+1}|h_k)$ est la probabilité que l'estimation de la position à l'instant $k+1$ soit correcte c'est-à-dire que l'amer correct ait été détecté (association de donnée correcte).
– $P(h_k)$ est un état caché du réseau bayésien dynamique. Il correspond à la probabilité $P(h_k|y_k)$ calculée par la précédente inférence bayésienne.

Le dernier terme à calculer est $P(h_{k+1}|h_k)$. L'étude d'un exemple permet d'illustrer le principe du calcul de cette probabilité (Figure 5.26).

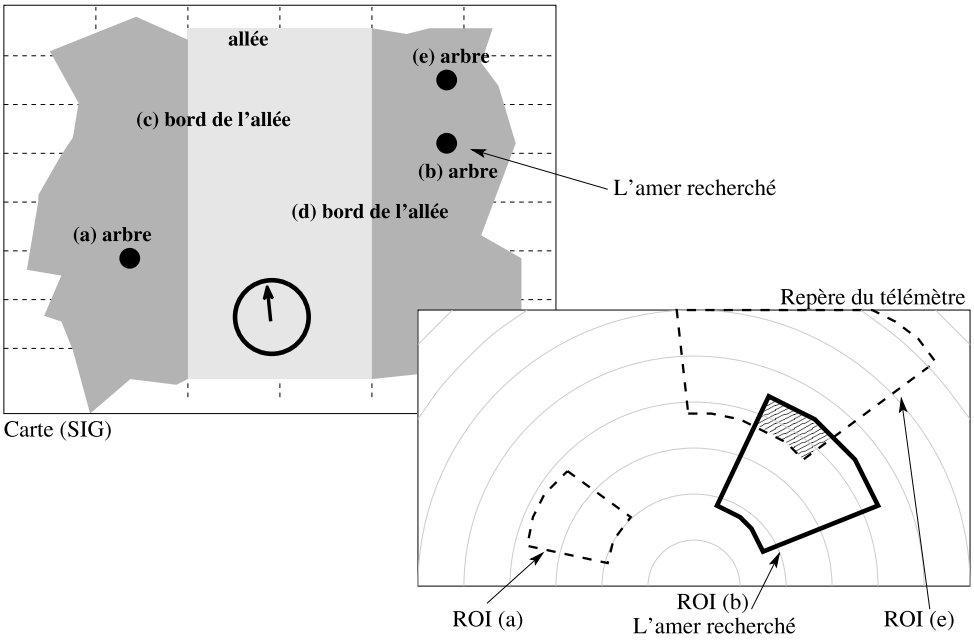


FIG. 5.26 – Intersection des zones d'intérêt. A gauche une vue de dessus de la carte avec les amers. Un triplet perceptif doit détecter l'amer (b) dans les données LIDAR. Sur la droite est tracé le polygone correspondant à la zone d'intérêt de ce triplet. Dans ces données LIDAR apparaissent d'autres amers détectables et leur zone d'intérêt tracée en pointillés.

Une vue de dessus de la carte est représentée à gauche. Un triplet perceptif doit détecter, l'amer (b) en le cherchant dans les données télémétriques issues d'un LIDAR. A droite est représentée la zone d'intérêt (polygone en trait continu) de ce triplet dans l'image LIDAR. Dans cet exemple, d'autres amers associés à leur triplet doivent être détectés. Par exemple l'un doit détecter l'amer (e) et un autre l'amer (a) dans cette même image LIDAR. Leurs zones d'intérêt sont représentées en pointillés. Deux zones d'intérêt se recouvrent. L'amer (e) est peut être dans la zone hachurée. Dans ce cas particulier, le triplet en charge de la détection de l'amer (b) dans la zone d'intérêt en trait continu court le risque d'une mauvaise association de données s'il détecte l'amer (e), le prenant pour l'amer (b). L'idée pour calculer $P(h_{k+1}|h_k)$ est de prendre en compte le fait que d'autres amers du même type que l'amer recherché peuvent appartenir à la zone d'intérêt de cet amer. En outre, plus la taille de la zone d'intérêt est importante, plus grande est la probabilité que du bruit soit détecté. Cette probabilité est divisée en deux termes :

$$P(h_{k+1}|h_k) = P(h_{k+1}|z_k) \cdot P(z_k|y_k, h_k) \tag{5.20}$$

où

– $P(z_k|y_k, h_k)$ est la probabilité que le triplet retourne un amer enregistré (i.e. z_k) du même type que l'amer recherché. Ainsi, cette probabilité correspond à la capacité du triplet à résister au bruit.
– $P(h_{k+1}|z_k)$ est la probabilité que le triplet retourne l'amer recherché parmi tous les amers du même type appartenant à la zone d'intérêt. Cette probabilité peut être aisément calculée : c'est un problème de combinatoire.

La figure 5.27 résume cette dernière idée. Finalement, la probabilité $P(h_{k+1})$ est :

$$P(h_{k+1}) = P(h_{k+1}|z_k) \cdot P(z_k|y_k, h_k) \cdot P(h_k|y_k) \tag{5.21}$$

et chaque triplet perceptif est défini par ces probabilités :

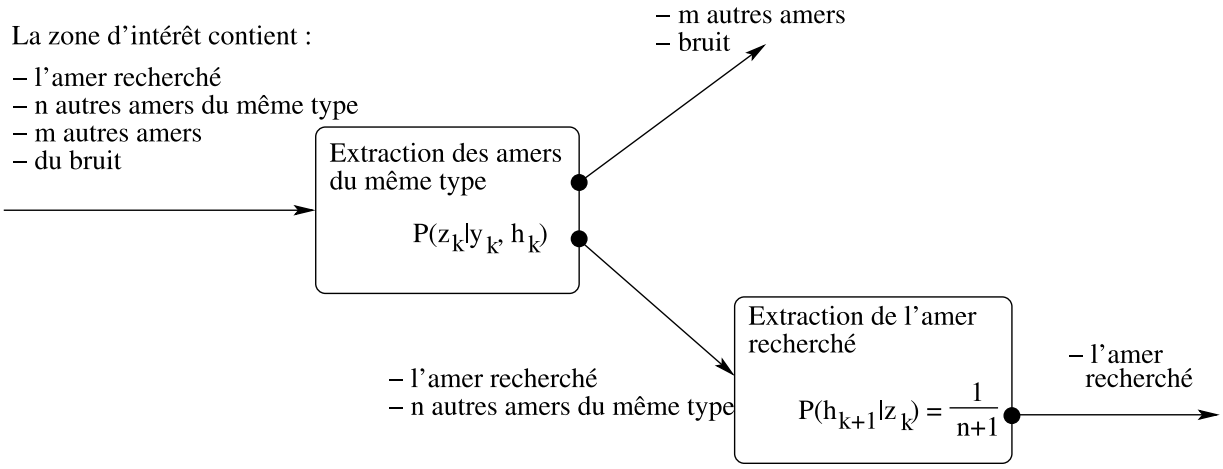


FIG. 5.27 – Principe du calcul de la probabilité de faire une association de données correcte.

– probabilité d'obtenir un résultat : $P(y_k|h_k)$ and $P(y_k|\overline{h}_k)$,
– probabilité de détecter un amer plutôt que du bruit : $P(z_k|y_k,h_k)$,
– probabilité de détecter l'amer recherché : $P(h_{k+1}|z_k)$.

Utilisation du niveau de confiance

L'utilisation du niveau de confiance est expliquée à partir d'un exemple (Fig. 5.28) qui compare l'approche proposée à une méthode plus classique de type MHT.

Le robot est supposé initialement situé dans un couloir. Cette position est représentée par l'état X_1 (diagramme 1.a sur la figure 5.28) associé à une confiance élevée (i.e. $P(h_1) = 0,84$) et à une précision médiocre. Le système est maintenant supposé capable de détecter des portes avec un capteur adapté. Puisque le robot se situe à proximité d'une porte, il crée un triplet perceptif chargé de détecter cette porte. Suite à la détection de cette porte dans une zone d'intérêt, l'état évolue vers une valeur X_2 (diagramme 1.b). Un niveau de confiance (par exemple $P(h_2) = 0,4$) est associé à ce nouvel état. Le robot se déplace alors. Le diagramme 1.c montre que l'état du robot est translaté dans la direction du déplacement. Il est plus dispersé ce qui est le reflet de l'incertitude apportée par le déplacement. Cependant, le niveau de confiance ne varie pas (§5.4.4). Dans cet état, le système sait grâce au SIG qu'une porte se situe à proximité. Un triplet est créé pour la détecter (diagramme 1.d). Grâce au système de focalisation, le triplet a peu de chance d'avoir fait une mauvaise association de donnée par conséquent le nouvel état (i.e. X_3) a une probabilité plus importante que le précédent ($P(h_3) = 0,56$). Le niveau de confiance de l'état est mis à jour : la dernière détection (diagramme 1.d) permet d'augmenter le niveau de confiance pour l'état X_2 : $P(h_2|y_2) = 0,73$.

Le compromis niveau de confiance vs. précision

L'objectif du système proposé est est d'offrir une estimation précise de la position (ellipsoïde d'incertitude de faible taille) et un haut niveau de confiance dans l'hypothèse de positionnement. Or la confiance et la précision sont des termes qui se contredisent. Dans le cas d'une estimation imprécise, la taille de la zone d'incertitude est importante et la probabilité que le robot soit situé effectivement dans cette zone est plus grande ce qui correspond à une confiance

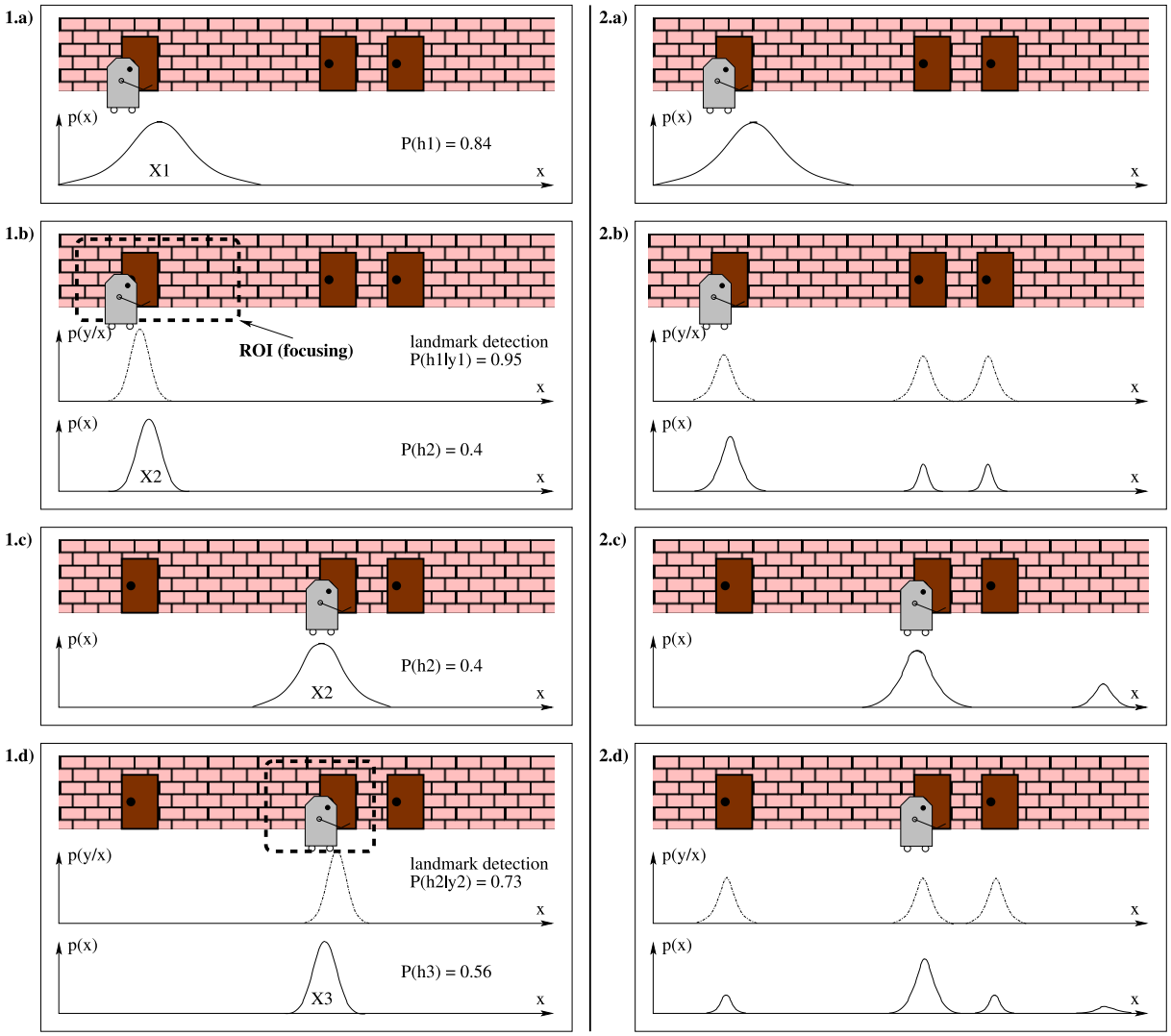


FIG. 5.28 – Comparaison de la méthode de localisation proposée (à gauche) et une approche classique basée sur une densité de probabilité multimodale (à droite).

élevée. Le système proposé n'échappe pas à cette règle. Le diagramme 1.b de la figure 5.28 montre un cas où le système a réussi la détection d'un amer. L'état X_2 qui en résulte est précis mais moins fiable que l'état initial ($P(h_2) = 0,4 < P(h_1) = 0,84$).

Le fait que la probabilité du nouvel état est calculée en prenant en compte la probabilité que l'amer effectivement détecté soit bien celui qui était recherché (Eq. 5.21) permet d'expliquer ce résultat. Initialement, l'état du véhicule est imprécis (état X_1) et la zone d'intérêt de l'amer recherché est de taille importante. Le triplet a donc un faible chance d'extraire l'amer correct ($P(z_k|y_k, h_k)$ est faible). De plus, comme il n'y a qu'un seul amer dans cette zone d'intérêt, $P(h_{k+1}|z_k) = 1$. Cependant, dans d'autres cas, cette probabilité peut être très faible. Ce scénario peu favorable se produit surtout en début de trajectoire alors que la localisation est peu précise. En fait quand l'estimation est précise, les zones d'intérêt sont petites et les probabilités $P(z_k|y_k, h_k)$ et $P(h_{k+1}|z_k)$ sont proches de 1. Dans ce cas, le niveau de confiance du nouvel état $P(h_{k+1})$ peut être supérieur à celui de l'état courant $P(h_k)$.

Mauvaise association de données

Une mauvaise association de donnée fausse systématiquement l'estimation. Les amers ne peuvent plus être détectés : les zones d'intérêt sont mal positionnées et elles ne les contiennent plus. Alors, grâce au système de focalisation, les détections échouent et la confiance diminue. Il suffit donc de surveiller le niveau de confiance pour détecter une mauvaise association de données.

Cependant, un faible niveau de confiance ne signifie pas forcément que le robot est mal localisé. En conséquence, la localisation est supposée fausse si le niveau de confiance est bas ET si une détection échoue. La dernière association de données est remise en cause et l'état du véhicule reprend sa dernière valeur valide. Il crée à partir de là un autre triplet pour améliorer la localisation.

5.4.5 Expérimentation et résultats

Généralités

Cette partie décrit les expérimentations réalisées pour valider l'approche. Un robot mobile tous terrains de type Robucar TT (fabriqué et commercialisé par la société Robosoft) est utilisé (Figure 5.29). Les capteurs qui y sont embarqués sont :

– un gyromètre à fibre optique KVH DSP 3000,
– des encodeurs sur les roues avant et arrière pour l'odométrie,
– un magnétomètre indique le nord,
– un récepteur GPS bas coût de précision $5m$,
– un télémètre LASER monoplan à balayage SICK PLS200 qui fournit des donnée de profondeur à la résolution de $0,5^o$ sur une dispersion angulaire de 180^o. Ces données sont disponibles toutes les $80ms$,
– une caméra CCD SONY VL500 qui fournit $7,5$ images couleurs par seconde de taille 640×480 pixels et avec l'encodage YUV422.

Tous ne sont pas utilisés pour cette application.

Le robot est également équipé de deux ordinateurs embarqués. Le premier utilisant Linux RTAI gère la commande bas niveau des moteurs du robot. Le second sous Linux, accueille le système de localisation et envoie des commandes de contrôle au premier.

Parmi l'ensemble des expérimentions réalisées, deux servent ici d'exemple. Dans la première les performances de localisation du système sont évaluées en statique (robot immobile). Dans la seconde l'approche proposée s'exécute dans une application réelle de guidage automatique en environnement extérieur.

Performances de l'algorithme

La figure 5.30 représente une vue de dessus de la carte de l'environnement expérimental. Cinq amers "arbres" y sont présents. Le cercle noir correspond à la position vraie du robot. La pose est initialisée grâce au GPS et au magnétomètre (cercle gris associé à l'ellipse d'incertitude en pointillés).

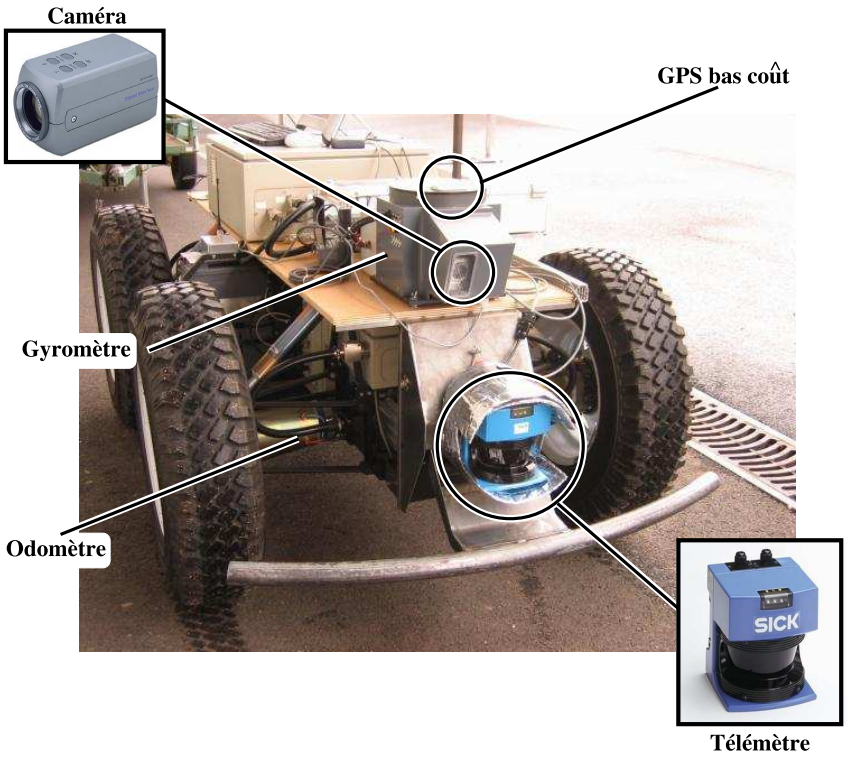


FIG. 5.29 – Robot mobile utilisé pour les expérimentations.

Cette première estimation est appelée ”Etat 0” sur la figure 5.31 qui montre les différents triplets perceptifs utilisés pour localiser le robot. Sur la partie gauche une vue de dessus de la carte présente le niveau de confiance (intégrité) de l'estimation.

La zone d'incertitude initiale qui caractérise l'Etat 0 correspond à l'ellipse à 3 écarts-types de la donnée GPS. Cette première estimation est supposée fiable (voir la probabilité associée à l'itération 0 sur la figure 5.32 qui montre le niveau de confiance associé à chaque état successif durant l'expérience).

A partir de cette première estimation, le système construit un ensemble de triplets. Dans le cas particulier de l'expérimentation présentée, tous les amers peuvent être détectés par le télémètre LASER. Ainsi, chaque triplet a en charge la détection d'un amer dans une zone d'intérêt des données télémétriques définie par le système de focalisation. Le système choisit alors le triplet 1 (figure 5.31). Ce triplet tente la détection de l'amer (a) ce qui réalise de fait l'association de donnée qui est malheureusement fausse dans le cas présenté. En conséquence, la position vraie est en dehors de la zone d'incertitude du robot. La probabilité de cet état est de $0,23$ (probabilité associé à l'itération 1 sur la figure 5.32). A ce stade, le système n'est pas encore capable de savoir que l'Etat 1 n'est pas valide. La zone d'intérêt de chaque triplet est alors mise à jour ce qui en réduit la taille. Comme l'estimation est invalide, ces nouvelles zones d'intérêt ne contiennent pas d'amer. En conséquence, le triplet 5 échoue dans la détection de l'amer (b). Cet échec fait chuter le niveau de confiance en dessous de $0,1$ (probabilité associée à l'itération 2 sur la figure 5.32). Cette non détection ET le bas niveau de confiance conduisent à la révision de la précédente association de données et le système revient dans l'Etat 0. Le système choisit alors le triplet 2 qui réussit la détection de l'amer (d) réalisant ainsi une association de données correcte. Finalement, tous les autres triplets détectent leur amer ce qui améliore la précision de localisation et augmente le niveau de confiance.

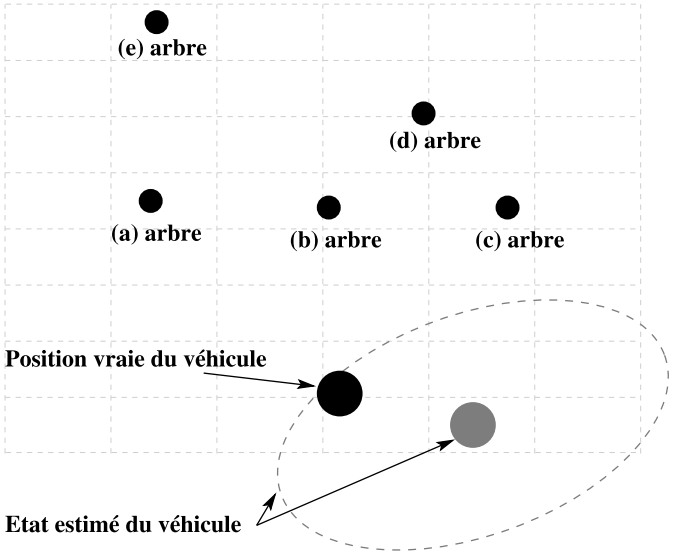


FIG. 5.30 – Vue de dessus de la carte avec 5 amers arbres, la position vraie du robot (cercle noir) et la pose initiale estimée (cercle gris) avec sa zone d'incertitude (ellipse en pointillés).

Application au guidage automatique du robot

Le but de l'application est le re-jeu autonome d'une trajectoire pré-définie à partir du système de localisation proposé. Cette trajectoire de référence d'une longueur de $300m$ est représentée figure 5.33.

De la même manière que précédemment, la pose du robot est initialisée à partir du GPS et du magnétomètre. A partir de cet état initial, le système est capable de détecter des amers de l'environnement en utilisant le système proposé. Le robot est ainsi localisé le long de la trajectoire à suivre.

Sept points de contrôle sont placés le long de la trajectoire et l'erreur de positionnement mesurée pour chacun d'eux. Cette définition "manuelle" des points de référence est rendue nécessaire par l'impossibilité de disposer d'une vérité de terrain sur l'ensemble de la trajectoire. La valeur correspondante des erreurs est donnée dans le tableau 5.2.

	A	B	C	D	E	F	G
moyenne/cm	-1.7	-0.2	0.5	-0.2	0	-0.7	-0.1
écart-type/cm	4.7	3.2	3.7	3.1	3.9	1.4	2.2
max/cm	9.2	6	7.3	4.5	6	3.6	5.2

TAB. 5.2 – Erreur de localisation aux 7 points de contrôle (de A à G) et sur 10 répétitions de la trajectoire de référence.

L'erreur maximale est inférieure à $10cm$. L'approche de détection active des amers dans un contexte multi-sensoriel permet de localiser le robot en satisfaisant les contraintes du temps réel.

A titre comparatif, la localisation du GPS seul durant le premier tour conduit à une erreur moyenne qui se situe entre $1m$ et $3m$. Il apparaît à l'évidence que l'utilisation du GPS seul ne permet pas une localisation du robot suffisamment fiable pour une navigation autonome.

Le système de localisation proposé fournit en sortie une estimation de la position, de sa

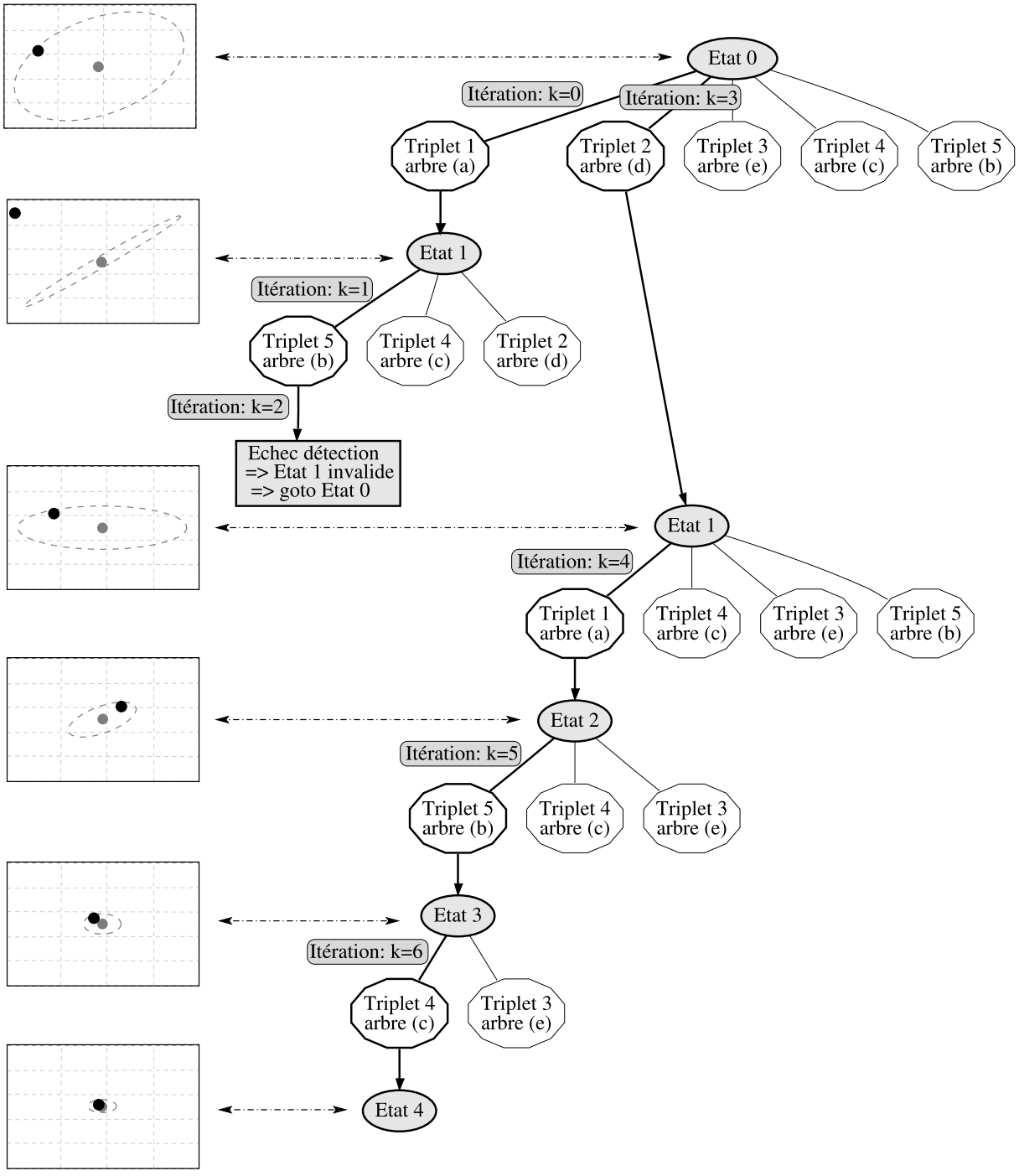


FIG. 5.31 – Localisation du robot à partir des triplet perceptifs. A gauche, une vue de dessus de la carte avec la pose vraie du robot (cercle noir) et l'état estimé (point gris et ellipse en pointillée).

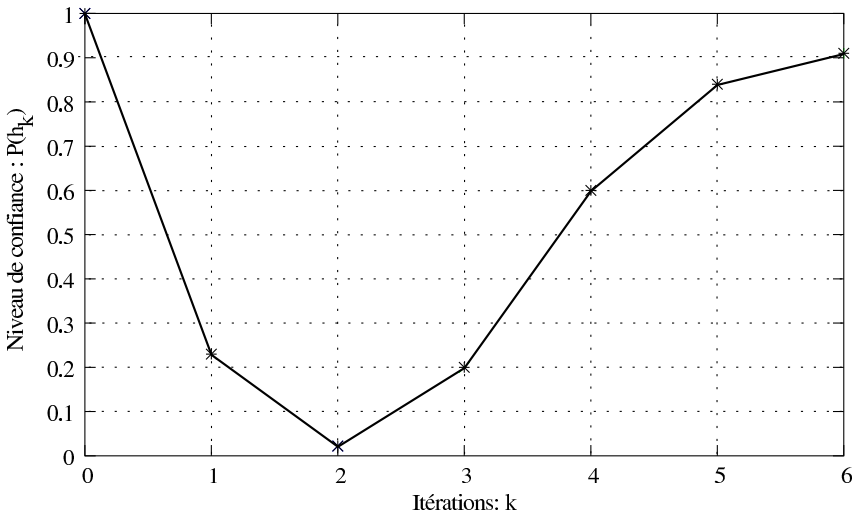


FIG. 5.32 – Evolution du niveau de confiance durant l'expérimentation.

FIG. 5.33 – La trajectoire de référence tracée sur une photo aérienne de la zone d'expérimentation.

précision et de son niveau de confiance. La figure 5.34 montre l'évolution du niveau de confiance de l'estimation et celle de la précision latérale (distance perpendiculaire entre la position courante et la trajectoire de référence) durant le premier tour.

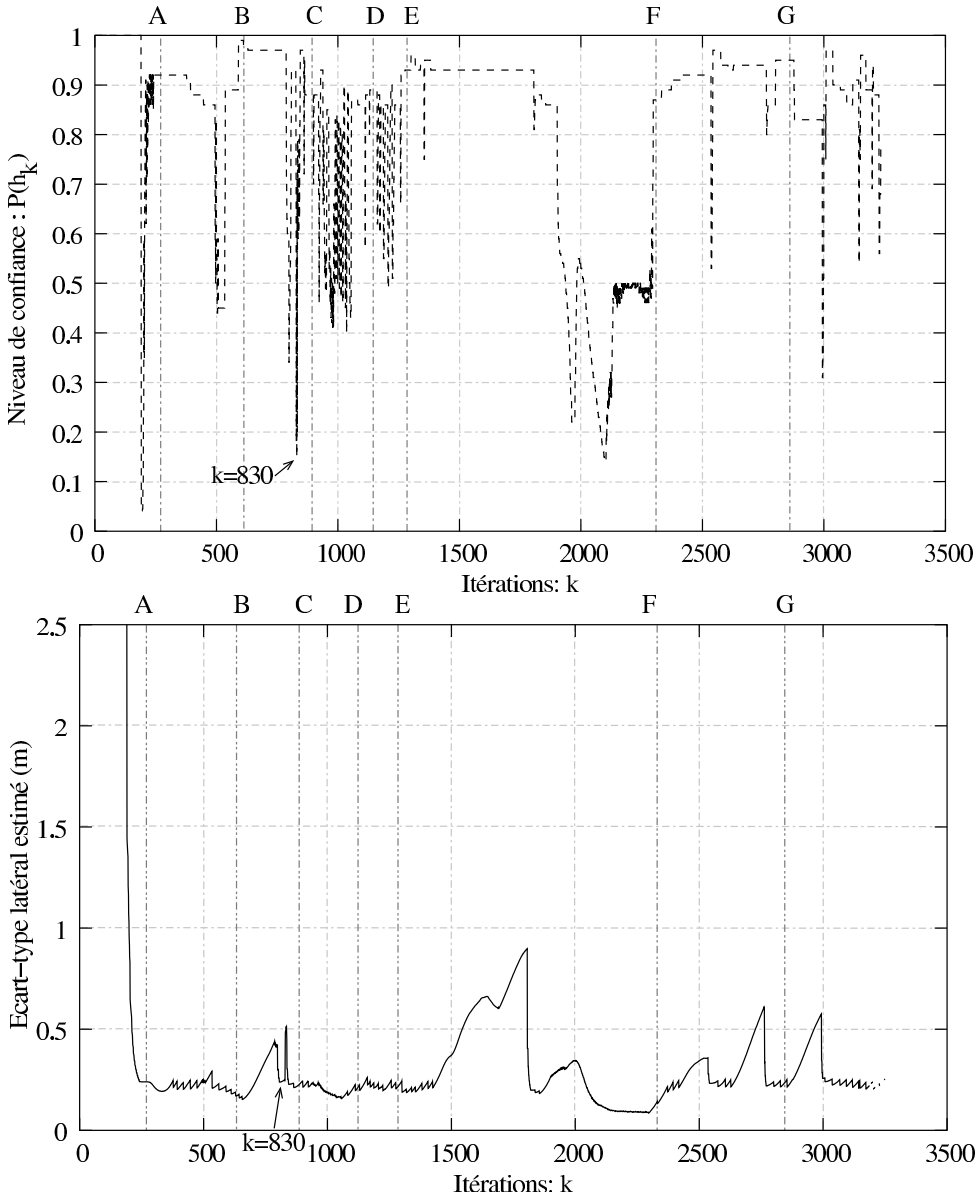


FIG. 5.34 – Evolution de la précision latérale et du niveau de confiance pendant le premier tour

Le système est initialement fiable ($P(h_0) = 1$ sur la figure 5.34) et imprécis (écart d'environ $2m$).

Comme l'objectif est d'obtenir une localisation précise et fiable le système génère des triplets perceptifs pour augmenter la précision. La première détection d'un amer à l'itération $k = 200$ permet d'améliorer la précision latérale ($0,25cm$) et de faire chuter le niveau de confiance à $0,05$. Cette chute importante du niveau de confiance est expliquée précédemment (§ 5.4.4).

Vers l'itération $k = 1500$, le niveau de confiance reste constant. En effet, comme aucun amer n'est visible dans cette zone, le robot navigue à l'estime à partir des capteurs proprioceptifs (codeurs et gyromètre) et dans ce cas, la confiance n'évolue pas (§ 5.4.4). Comme aucune détection d'amer n'améliore la précision, la précision diminue (§ 5.4.4).

Entre les points de contrôle C et E, la localisation est précise en raison de détections d'amers plus nombreuses.

Un mauvaise association de données se produit vers l'itération $k = 380$. Le niveau de confiance est très bas ($P(h_{830}) = 0,16$) et la précision latérale est faible ($0,25cm$). La dernière association de donnée est remise en cause et la confiance augmente ($P(h_{831}) > 0,6$) ainsi que l'erreur de localisation latérale ($0,52cm$), (Figure 5.35).

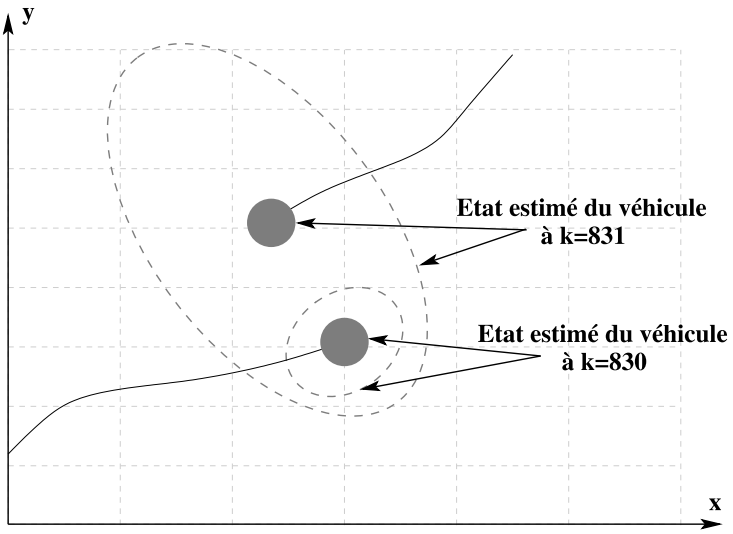


FIG. 5.35 – Vue de dessus de la carte. A $k = 830$, l'état du robot est invalide. A $k = 831$, le système détecte l'invalidité, annule la mauvaise association de données pour retrouver un état valide mais moins précis à $k = 831$.

5.4.6 Conclusion et perspectives

Cette partie présente un système de localisation et de guidage d'un robot mobile en environnement extérieur. Il est basé sur une approche multi-sensorielle et sur l'utilisation d'un processus de focalisation attentionnelle des détections d'amer qui sont réalisées par des entités autonomes appelées "triplets perceptifs". Le principe de focalisation améliore la détection des amers. De plus, il permet de mettre en place un système de réaction aux mauvaise associations de données. La non détection d'un amer dans la zone d'intérêt où il est attendu témoigne d'une probable mauvaise localisation du robot due à un mauvais appariement. Ce processus, placé dans un cadre probabiliste, permet de calculer le niveau de confiance de l'estimation courante de la position. Cette approche a été validée expérimentalement par une application de guidage automatique d'un robot en situation réelle.

Une évolution envisagée consiste à améliorer le critère de choix des triplets perceptifs en le basant sur une expression plus précise des buts à atteindre pour une application donnée. De plus, les triplets sont traités un par un. L'implémentation de cette méthode sur une architecture parallèle permettrait de suivre plusieurs hypothèses (i.e. exécution de plusieurs triplets) en même temps et la coopération entre eux permettrait de réduire encore les ambiguïtés dans l'association de données ce qui améliorerait la confiance. Cette adaptation à un environnement de calcul parallèle du système de perception attentionnel par focalisation progressive constitue le sujet de la thèse de David Grimbichler qui a débuté en 2006 et que je co-encadre à 50%.

6 Perception des objets dynamiques

6.1 Introduction

L'utilisation d'une carte comme modèle de l'environnement pour guider la reconnaissance des amers dans le cadre d'une application de localisation est réaliste : l'obtention préalable d'une carte numérique des amers est possible (et le sera de plus en plus) car ils constituent des points de repère fixes et à durée de présence suffisante. Mais la perception d'une scène en vue d'applications robotiques va plus loin que le problème de la localisation. Il est nécessaire de traiter aussi les éléments dynamiques (autres robots, piétons, obstacles divers) qui présentent, du point de vue de leur perception, deux particularités :

– Ils peuvent être ou ne pas être présents dans la scène. Cet aspect dynamique interdit qu'une carte immuable des obstacles puisse être utilisée pour guider la perception. Les obstacles détectés sont rajoutés temporairement dans la carte par le processus de perception lui même : ils sont présents dans la carte tant qu'ils sont visibles. Un processus de perception de type *bottom-up* est mis en oeuvre pour rechercher de nouvelles occurrences d'objets mobiles qui sont aussitôt ajoutés à la carte. Il peuvent alors être utilisés par le processus *top-down* comme des amers particuliers qui guident la perception en permettant la focalisation de leur recherche dans des zones d'intérêts des données capteurs.
– Ils présentent le plus souvent une grande variabilité d'apparence qui rend leur modélisation délicate. Or si une représentation de ces éléments dynamiques de la scène n'est pas disponible, le processus de perception ne peut être guidé et le système attentionnel n'est pas mis en œuvre. Dans ce cas, le modèle des objets à reconnaître est obtenu par apprentissage de leur apparence locale grâce à une base de données annotées. Ce modèle appris guide ensuite la perception.

Chacune de ses deux problématiques fait l'objet d'une étude particulière qui est présentée dans les deux prochaines sections.

6.2 Reconnaissance des objets par vision focalisée

6.2.1 Introduction

Le système de perception attentionnelle est ici utilisé pour apporter une contribution originale au problème de la reconnaissance et du suivi d'objet dans des images. Cette application correspond aux travaux de thèse de Noël Trujillo soutenue le 13 décembre 2007 [63]. Elle constitue la première étape de l'utilisation de la méthode vers l'analyse de scène, c'est-à-dire vers la reconnaissance des différents objets qui la composent en tenant compte des relations statistiques entre ces objets. Dans le cas de la reconnaissance d'objet, il s'agit de considérer que l'objet à reconnaître constitue l'objet racine dans la structure hiérarchique (figure 4.2 au §4.3.1). Ces objets fils sont des parties discriminantes de l'objet qui peuvent être chacune reliée à des primitives de natures diverses observables dans l'image par l'intermédiaire de leur fonction de détection.

Dans ce contexte, le système est mis en œuvre dans un cadre monocapteur (caméra vidéo). Ce travail intègre dans un même système de perception un certain nombre de points importants de la reconnaissance d'objet comme l'analyse multi-caractéristiques ou l'analyse multi-échelle. L'apprentissage du modèle qui supervise la reconnaissance utilise un ensemble d'images annotées contenant une instance de la classe d'objet à reconnaître.

Le problème de la sélection des caractéristiques (réduction de la dimension de l'espace des paramètres appris) est résolu par analyse de la matrice de covariance des paramètres appris. Ainsi non seulement les caractéristiques les plus discriminantes (c'est-à-dire celles dont les paramètres ont une variance faible) sont retenues mais aussi celles qui statistiquement très corrélées avec une autre, forment une paire discriminante.

6.2.2 Focalisation de l'attention en reconnaissance d'objets

Grâce au développement et à la mise au point de techniques de classification (réseaux de neurones, SVM[1]), et d'outils de traitement d'images, la technique du template matching a été largement utilisée. Les approches appartenant à cette catégorie balayent l'image de manière exhaustive en position et en échelle afin de reconnaître l'objet [65, 73, 74]. L'utilisation de classifieurs en cascade [83, 73], l'introduction d'indices de focalisation comme la couleur pour les visages ou les ombres pour la détection de véhicules, ou la délimitation d'une zone d'intérêt dans l'image définie par l'utilisateur [65], sont utilisés comme des processus simples de focalisation attentionelle pour réduire le temps de calcul. En outre, la réduction du nombre de paramètres dans le vecteur des caractéristiques [65], la réduction du nombre d'échelles ou du pas de résolution spatiale sont des étapes difficiles. Il en résulte une augmentation du nombre de fausses détections (classifieur peu discriminant) ou de mauvaise détection pour les objets variables en taille.

D'autres approches appartenant à la catégorie dite d'appariement relationnel existent. Elles sont bien adaptées pour reconnaître des objets déformables ou d'apparence variable. Dans ce cas, et à la différence des approches purement structurelles, il n'est pas difficile d'apprendre l'apparence de chaque partie composant l'objet (patchs). En revanche, elles doivent être définies par l'utilisateur, comme c'est le cas dans [51, 16, 56, 31]. Du fait de la mise en

[1] Suport Vector Machine

correspondance, le problème de la combinatoire se pose là aussi. Pour en diminuer l'impact, l'espace de recherche peut être limité en introduisant des contraintes sur la configuration spatiale [51].

Afin d'optimiser le processus de perception, plusieurs mécanismes attentionnels ont été proposés. Des approches *bottom-up* [80, 77, 42, 84], ou *top-down* [85, 75, 36, 52, 58, 69] montrent de bonnes performances pour la sélection de régions d'intérêt. Quand il s'agit de reconnaître un objet, la stratégie classique est la suivante : un module d'attention visuelle est chargé d'orienter un module de classification vers des régions d'intérêt [26, 42, 4, 25, 85, 58, 69, 36]. Ces deux modules sont considérés comme indépendants.

Au bilan, la reconnaissance d'objets fait apparaître d'une part le besoin d'orienter les modules de reconnaissance d'objets au moyen d'un mécanisme attentionnel et d'autre part la tendance des approches du type *template matching* à décomposer l'objet en un ensemble de parties pour d'accélérer la reconnaissance.

6.2.3 Vision focalisée

L'objectif est de développer une structure où la reconnaissance d'objets (vue comme un processus de recherche visuelle) et l'attention visuelle sont gouvernées par le même algorithme de contrôle. Ceci nécessite une représentation particulière de l'objet. Dans cette approche, l'attention sélective opère sur quatre niveaux différents :
– sélection des opérateurs bas niveau,
– intervalle d'action des opérateurs choisis (sous-espaces de l'espace des caractéristiques),
– niveau(x) de résolution sur le(s)quel(s) chaque opérateur doit agir,
– région de l'image dans laquelle ils doivent agir (fenêtrage).

Représentation de l'objet

Un objet peut être défini comme une entité indépendante caractérisée par différentes propriétés telles que : la couleur, la taille, la forme, la texture, etc. Pour la tâche de reconnaissance d'objets, la question de savoir quelle représentation est la plus pertinente se pose d'emblée non seulement pour permettre leur reconnaissance mais aussi pour pouvoir être exploitée dans un cadre de la *vision guidée par l'objectif*. Le type de représentation doit être adapté à la tâche à réaliser. La définition d'un modèle général pour un objet quelconque nécessite de répondre aux questions suivantes :

– Quelle propriété représente le mieux l'objet ?
– A quelle position et échelle d'analyse telle ou telle propriété doit-elle être observée ?
– Comment faire pour éviter d'indiquer *explicitement* les régions ou composantes supposées pertinentes pour la bonne caractérisation de l'objet ?

L'idée de départ est de choisir une représentation hybride de l'objet : une représentation basée sur l'apparence mais qui intègre aussi un ensemble de parties composant la structure de l'objet. Cette idée n'est pas nouvelle. Schneiderman [73] propose une représentation du même type, dans laquelle l'objet est représenté par un ensemble de parties ; chacune correspondant à une région dans l'image (groupe de pixels) qui modélise la structure locale, tout en gardant les dépendances statistiques des parties de l'objet. Avec la notion de partie, l'objet n'est pas seulement représenté *a priori* par des segments, des surfaces ou des coins etc., ou par un

modèle de connaissance de l'objet mais plutôt par des représentations acquises et apprises par le système lui-même à partir des stimuli (sorties des opérateurs bas niveau).

La caractérisation d'une partie par la statistique sur sa structure locale facilite sa détection. Dans le cadre du système attentionnel, les corrélations statistiques entre les différentes parties complètent cette caractérisation de telle manière que la détection d'une partie contraigne celles des autres parties. Ceci constitue l'adaptation au cas de la reconnaissance d'objet du principe focalisant sur lequel repose le système de perception proposé.

Constituants d'un objet : définition d'une partie

Dans la méthode proposée, l'objet est considéré comme composé d'un ensemble de **parties** dépendantes entre elles. Chaque partie correspond à une région de l'image (groupe de pixels) laquelle décrit la structure locale de l'objet. Elle est désignée par Λ_m pour $m \in [1, M^p]$, où M^p est le nombre total de parties constituant l'objet.

Pour une partie Λ_m quelconque, sont définis :

- **Un vecteur de paramètres,** λ_m défini par $\lambda_m = [\zeta_{m1}, \zeta_{m2}, \ldots, \zeta_{mN_m}, \mathbf{a}_m^t]^t$. Les éléments ζ_{ij} peuvent-être : une propriété physique comme la couleur ou un niveau de gris moyen, la réponse fréquentielle dans une zone donnée de l'image, etc. . .$\mathbf{a}_m = [u_m, v_m]^t$ désigne les coordonnées de Λ_m dans l'image. Cette représentation multi-caractéristiques est de première importance dans le processus de focalisation.
- **Un poids** w_m qui indique la pertinence dans le processus de reconnaissance, de la partie à laquelle il est associé. Ce poids est important car il permet d'optimiser le processus de perception en sélectionnant, par exemple, la partie qui a une probabilité de présence élevée.
- **Un indice** r_m qui sert à indiquer le niveau de résolution auquel appartient la partie Λ_m.
- **Une fonction de détection** $\hat{\lambda}_m = f_m(\bar{\lambda}_m, C_{\lambda_m})$. Associée à une partie Λ_m elle extrait une observation $\hat{\lambda}_m$ dans une région d'intérêt centrée sur $\bar{\lambda}_m$, où $\bar{\lambda}_m$ et C_{λ_m} correspondent respectivement au vecteur de paramètres moyen et à sa matrice de covariance. Cette région d'intérêt est définie dans l'espace des caractéristiques de dimension $(N_m + 2)$ et non pas seulement dans l'espace géométrique.

Une partie correspond donc dans le cas de cette application à la notion d'objet définie dans la description générale du système de perception attentionnelle par focalisation progressive présenté à la section 4.3.1. Comme l'objet générique, la partie est associée à un vecteur de paramètres, à une fonction d'utilité (ici le poids w_m) et à une fonction de localisation (ici la fonction de détection).

Afin de savoir à quelle région de l'image correspond une partie Λ_m, une grille de M^c cellules définie ci-après est utilisée. Cette grille a pour rôle de disséminer les paramètres dans plusieurs régions de l'image qui pourraient, lors de l'apprentissage, être définies comme des *parties* représentatives de l'objet.

Définition d'une cellule

Une cellule $\mathcal{C}$, est définie comme une entité qui extrait la structure locale d'une image à partir d'opérateurs bas niveau. Cette entité présente l'intérêt de pouvoir concentrer de multiples propriétés de l'objet dans une région spécifique de l'image. A l'intérieur de chaque cellule, les

donnés provenant des N ($N > N_m$) opérateurs bas niveau définis avant l'apprentissage, sont accessibles. Grâce à l'apprentissage, ce type de représentation permet focaliser les détections à partir d'indices comme par exemple la couleur, la direction du gradient dans une partie de l'image,...

Une cellule $\mathcal{C}$, se différencie d'une partie en ce sens qu'elle sert à **extraire** l'information locale d'une image à partir des opérateurs alors qu'une *partie* correspond à une **cellule caractérisée** lors de l'apprentissage par la statistique locale de l'objet. Elle devient alors un descripteur local.

Pour une cellule particulière $\mathcal{C}_m$ d'indice m et de vecteur de paramètres $\varsigma_m = [\zeta_{m1}, \zeta_{m2}, \ldots, \zeta_{mN}, \mathbf{a}_m^t]^t$, les N opérateurs sont supposés statistiquement indépendants et la matrice de covariance C_{ς_m} est diagonale.

En revanche, la dépendance entre plusieurs cellules pour un même opérateur est conservée. Les $[\sigma^2_{\zeta_{m1}}, \sigma^2_{\zeta_{m2}}, \ldots, \sigma^2_{\zeta_{mN}}, C_{\mathbf{a}_m}]$ correspondent aux éléments de la diagonale principale. N est le nombre total des opérateurs bas niveau, $\mathbf{a}_m$ et $C_{\mathbf{a}_m}$ correspondent respectivement au vecteur des coordonnées de $\mathcal{C}_m$ et à sa matrice de covariance.

La non connaissance a priori de la signification de chaque partie de l'objet, impose de distribuer les cellules de façon uniforme selon une grille carrée ou rectangulaire (en fonction de la boîte englobante de l'objet à apprendre ou de la constitution de la base de données). Cette grille est placée sur l'objet à reconnaître qui doit être centré. C'est pourquoi, une base de données annotée ou bien des images avec des objets plus ou moins centrés et normalisés en taille, sont nécessaires afin de placer correctement la grille sur l'objet.

Afin de disposer d'une représentation multi-échelle pour avoir une représentation plus complète de l'objet, une grille multi-résolution est nécessaire. Une intégration de ce type présente l'avantage de permettre au processus de reconnaissance d'avoir des indices multi-échelle pour la focalisation. De nombreux travaux ont utilisé une approche coarse to fine (de l'échelle grossière à l'échelle plus fine) afin d'utiliser des descripteurs à basse résolution (échelle grossière) pour détecter plus rapidement des descripteurs à haute résolution. Le choix de l'échelle d'analyse n'est pas défini a priori : elle dépend des besoins et de l'état courant du processus de reconnaissance attentionnelle. Pour la grille de cellules multi-résolution une analyse multi-résolution typique 2^n est réalisée, ce qui signifie que l'image est décomposée en n résolutions dont le sous-échantillonnage est une puissance de 2. Le nombre total de cellules M_c est alors la somme du nombre de cellules M_i à chacune des N_r résolutions : $M_c = \Sigma_{i=1}^{N_r} M_i$. La grille multi-résolution utilisée est présentée figure 6.1.

6.2.4 Apprentissage du modèle

Le but ici est de fournir un modèle initial de l'objet en question, c'est-à-dire de définir chaque partie de l'objet ; il s'agit de donner une valeur moyenne et une dispersion à chaque paramètre de chaque cellule. Ainsi, de la grille de M^c cellules, sont extraites les M^p cellules les plus caractéristiques de l'objet ainsi que les N_m paramètres les plus discriminants de celles-ci (N_m est potentiellement différent pour chaque cellule, et $N_m \leq N$). Les M^p cellules sélectionnées sont les *parties* composant l'objet.

L'apprentissage est divisé principalement en trois étapes : l'apprentissage des paramètres, l'apprentissage sur le positionnement des cellules et la sélection des caractéristiques.

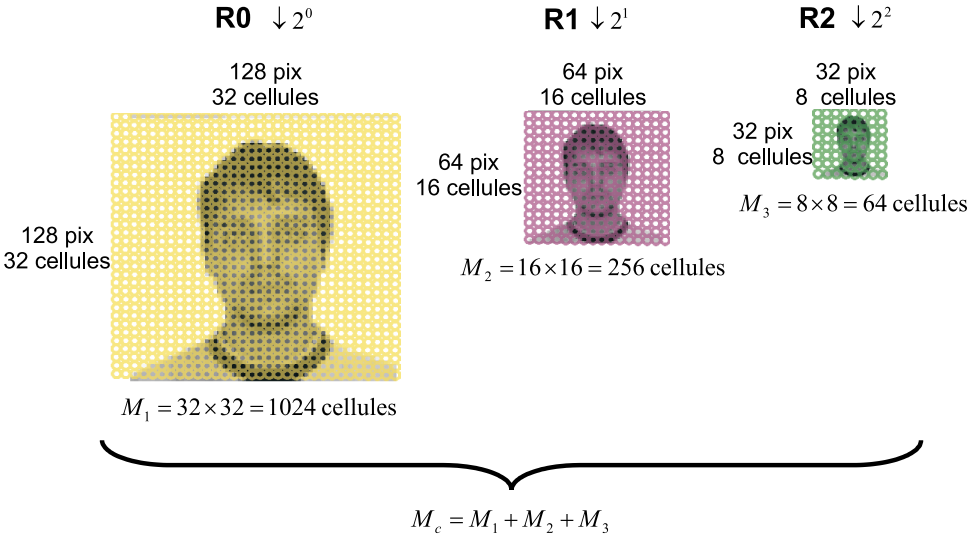


FIG. 6.1 – Exemple d'une grille tri-résolution.

Apprentissage des paramètres

Il s'agit de déterminer, pour chaque cellule $\mathcal{C}_m$ et à partir de T exemples, la valeur moyenne et la dispersion du vecteur de paramètres ς_m : $\overline{\varsigma}_m$ et C_{ς_m}.

L'apprentissage débute par la collecte des statistiques de chaque cellule : les N opérateurs Op_n, pour $n \in [1, N]$, disposés au départ pour l'apprentissage fournissent une valeur ζ_{mn}.

Pour un exemple donné, les N opérateurs sont appliqués et leur réponse à l'intérieur de chaque cellule est calculée. L'hypothèse d'indépendance statistique entre les opérateurs pour une cellule donnée (la dépendance entre plusieurs cellules est conservée), permet le calcul de chaque opérateur Op_n individuellement. Ce processus est répété pour chacun des T exemples ; N matrices, $\mathbf{OP}_n = [\mathbf{op}_{n1}\mathbf{op}_{n2}\ldots\mathbf{op}_{nT}]$ ($n \in [1, N]$), de taille $M^c \times T$ sont calculées. M^c est le nombre total de cellules et T est le nombre total d'exemples.

Positionnement initial des cellules

Puisqu'un objet quelconque peut être situé à n'importe quel endroit de l'image, à différentes échelles et avec des petites variations de point de vue (rotation 2D et 3D), il faut que le modèle de l'objet intègre ces variations possibles pour avoir une reconnaissance invariante en position, échelle et point de vue. L'ensemble des exemples d'apprentissage peut se présenter sous deux formes

1. Une base de donnée d'images annotées (c'est-à-dire des images où un objet est présent et où sa position est connue et répertoriée). Dans ce cas, il est possible d'apprendre directement la position de chaque cellule.

2. Une base d'image ou l'objet est centré et normalisé en taille. Dans ce cas, la position possible des cellules est apprise par une transformation (similitude 2D) de l'image d'apprentissage normalisée.

Sélection des caractéristiques

Lors de la caractérisation des cellules à partir des exemples, il apparaît que certaines ne sont pas positionnées sur l'objet mais sur le fond ou plus généralement sur une région de l'image qui n'est pas du tout descriptive de la classe d'objets en question. Toutes ces cellules ne sont pas utiles et peuvent être éliminées. Cette élimination des cellules n'est cependant pas la seule possibilité pour simplifier le modèle. Il est aussi possible de réduire la dimension de l'espace des caractéristiques en éliminant des paramètres qui ne sont pas descriptifs d'une région échantillonnée par une cellule quelconque. Deux techniques sont donc décrites : l'élimination des paramètres non descriptifs et l'élimination des cellules non fonctionnelles.

– **Elimination des paramètres non descriptifs**
Lors de la collecte des statistiques pour une cellule donnée, il y a des paramètres qui ne décrivent pas l'objet. Ce manque de description peut être dû à deux raisons : soit l'information appartient au fond ou à une région qui n'est pas descriptive de la classe d'objets, soit l'opérateur bas-niveau ne fournit pas de réponse (exemple calcul de la direction du contour sur une région où il n'y a pas de contour). En effet, la *connaissance* qu'a le système d'un objet déterminé est donnée par la matrice de covariance et par la statistique sur la réponse des opérateurs qui sont utilisées pour sélectionner l'information nécessaire au guidage du processus de reconnaissance.
Deux critères de sélection extraits de la matrice de covariance des paramètres des cellules sont utilisés :
  – **la variance** d'un paramètre pour une cellule déterminée : si sa valeur est élevée c'est que ce paramètre est statistiquement très dispersé pour la classe d'objet et donc qu'il n'est pas discriminent.
  – **la corrélation** d'un même paramètre dans un cellule donnée avec ce même paramètre dans une ou plusieurs autres cellules : le fait que cette corrélation soit très élevée s'avère être une caractéristique discriminante de l'objet. Autrement dit, un objet est très bien caractérisé par le fait que si une caractéristique est présente alors une autre l'est toujours, même si, prises individuellement, ces caractéristiques sont peu discriminantes. Un bon choix est de garder un paramètre dans une cellule s'il est corrélé avec au moins une autre cellule avec un coefficient de corrélation supérieur à un seuil, 95% par exemple.

  Pour faciliter l'élimination des paramètres non discriminants pour une cellule déterminée, il convient de représenter le modèle sous la forme suivante (les opérateurs sont supposés décorrélés) :
  $\mathbf{x}^p = [\bar{\mathbf{op}}_1, \bar{\mathbf{op}}_2, \ldots, \bar{\mathbf{op}}_N, \bar{\mathbf{a}}]^t$ et la matrice de covariance diagonale $\mathsf{C}_{\mathbf{x}^p}$ où $[\mathsf{C}_{\mathbf{op}_1}, \mathsf{C}_{\mathbf{op}_2}, \ldots, \mathsf{C}_{\mathbf{op}_N}, \mathsf{C}_{\mathbf{a}}]$ correspondent aux éléments de la diagonale principale ; avec

$$\bar{\mathbf{a}} = \begin{bmatrix} \bar{\mathbf{a}_1} \\ \bar{\mathbf{a}_2} \\ \vdots \\ \bar{\mathbf{a}_{M^c}} \end{bmatrix} \text{ et } \mathsf{C}_{\mathbf{a}} = \begin{pmatrix} \mathsf{C}_{\mathbf{a}_1} & \cdots & \\ & \mathsf{C}_{\mathbf{a}_2} & \\ \vdots & & \ddots \\ & & \mathsf{C}_{\mathbf{a}_{M^c}} \end{pmatrix}$$

  Pour chaque opérateur Op_n, la matrice corrélation est calculée à partir de la matrice de covariance, par la relation suivante : $r_{\zeta_i\zeta_j} = \frac{\sigma_{\zeta_i\zeta_j}}{\sqrt{\sigma_{\zeta_i\zeta_i}\sigma_{\zeta_j\zeta_j}}}$, avec $-1 \leq r_{ij} \leq 1$.
  Les paramètres éliminés sont ceux qui obéissent à $|r_{ij}| < 0.95$. Leur influence dans certaines cellules est ainsi supprimée.
  A titre d'exemple, si $\mathbf{op}_n = [\zeta_{1n}, \zeta_{2n}, \zeta_{3n}, \ldots, \zeta_{M_n^c}]$, seuls les M^p paramètres restants $\mathbf{op}_n^p = [\zeta_{1n}, \cancel{\zeta_{2n}}, \cancel{\zeta_{3n}}, \ldots, \zeta_{M_n^c}]^t$ sont gardés (e barré indique les paramètres éliminés).
  Donc, après élimination le modèle est réduit :

$$\mathbf{x}_r^p \sim \begin{cases} \mathcal{N}(\overline{\mathbf{op}}_1^p, \mathsf{C}_{\mathbf{op}_1^p}), \\ \mathcal{N}(\overline{\mathbf{op}}_2^p, \mathsf{C}_{\mathbf{op}_2^p}), \\ \vdots \\ \mathcal{N}(\overline{\mathbf{op}}_N^p, \mathsf{C}_{\mathbf{op}_N^p}), \\ \mathcal{N}(\overline{\mathbf{a}}, \mathsf{C}_{\mathbf{a}}). \end{cases}$$

où $\overline{\mathbf{op}}_n^p$ est le vecteur réduit de taille $M_n \leq M^c$. La matrice de covariance $\mathsf{C}_{\mathbf{op}_n^p}$ a une taille $M_n \times M_n$. L'élimination d'un paramètre du vecteur $\overline{\mathbf{op}}_n^p$ revient à ne pas faire l'observation de ce paramètre dans la cellule correspondante.
Concernant le traitement bas-niveau, dans le cas où le paramètre ζ_{in} a été éliminé, le module correspondant à l'opérateur Op_n est inhibé ce qui revient à faire la *focalisation dans l'espace des opérateurs*.

– **Modèle après élimination des paramètres**
Pour chaque cellule, le vecteur moyen s'écrit :
$\overline{\varsigma}_m^p = [\overline{\zeta'}_{m1}, \overline{\zeta'}_{m2}, \ldots, \overline{\zeta'}_{mN_m}, \overline{\mathbf{a}'}_m]^t$
et la matrice de covariance diagonale est :
$\mathsf{C}_{\varsigma_m^p} = [\sigma^2_{\zeta'_{m1}}, \sigma^2_{\zeta'_{m2}}, \ldots, \sigma^2_{\zeta'_{mN_m}}, \mathsf{C}_{\mathbf{a}'_m}]$.
Le modèle devient alors :

$$\mathbf{x}_r = \begin{bmatrix} \varsigma_1^p \\ \varsigma_2^p \\ \vdots \\ \varsigma_M^p \end{bmatrix}, \mathsf{C}_{\mathbf{x}_r} = \begin{pmatrix} \mathsf{C}_{\varsigma_1^p} & & \cdots \\ & \mathsf{C}_{\varsigma_2^p} & \\ & \ddots & \\ \vdots & & \mathsf{C}_{\varsigma_M^p} \end{pmatrix}$$

Il faut remarquer que, après la sélection des caractéristiques, chaque cellule a un nombre de paramètres $N_m \leq N$ et que $\mathsf{C}_{\mathbf{x}_r}$ n'est pas une matrice diagonale.

– **Elimination des cellules non fonctionnelles : obtention des parties**
Dès l'élimination des paramètres non descripteurs de l'objet à l'intérieur de certaines cellules, il peut y avoir des cellules sans aucun paramètre pertinent. Dans ce cas elles sont éliminées de la grille.
Après l'élimination des cellules non fonctionnelles, le modèle final réduit est

$$\mathbf{x}_f = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{M^p} \end{bmatrix}, \mathsf{C}_{\mathbf{x}_f} = \begin{pmatrix} \mathsf{C}_{\lambda_1} & & \cdots \\ & \mathsf{C}_{\lambda_2} & \\ & \ddots & \\ \vdots & & \mathsf{C}_{\lambda_{M^p}} \end{pmatrix}$$

avec $\lambda_m = [\zeta'_{m1}, \zeta'_{m2}, \ldots, \zeta'_{mN_m}, \mathbf{a}'_m]^t$ le vecteur moyen et C_{λ_m} la matrice de covariance.
$[\sigma^2_{\zeta'_{m1}}, \sigma^2_{\zeta'_{m2}}, \ddots, \sigma^2_{\zeta'_{mN_m}} \mathsf{C}_{\mathbf{a}'_m}]^t$ sont les éléments de la diagonale principale avec $m \in [1, M^p \leq M]$.
Le vecteur λ_m correspond au vecteur de paramètres de chaque partie Λ_m.

En résumé, sont disponibles après l'apprentissage :
– un modèle $X \sim \mathcal{N}(\overline{\mathbf{x}}, \mathsf{C}_{\mathbf{x}})_{obj}$ de l'objet contenant M^p *parties* Λ_m, avec $m \in [1, M^p \leq M^c]$, étant des cellules caractérisées par
  – un vecteur de paramètres $\lambda_m = [\zeta'_{m1}, \zeta'_{m2}, \ldots, \zeta'_{mN_m}, \mathbf{a}'_m]$, avec $N_m \leq N$,

– une matrice de covariance diagonale avec pour éléments de la diagonale principale $C_{\lambda_m} = (\sigma^2_{\zeta'_{m1}}, \sigma^2_{\zeta'_{m2}}, \ldots, \sigma^2_{\zeta'_{mN_m}}, C_{\mathbf{a}'_m})$ indiquant leur dispersion, où $C_{\mathbf{a}'_m}$ correspond à la matrice de covariance des coordonnées de la cellule m.

6.2.5 Stratégie de reconnaissance

Dans le contexte de la reconnaissance d'objet, l'utilisation d'un mécanisme de focalisation est évidemment motivé par le but d'optimiser le processus de reconnaissance. La mise en œuvre d'un tel processus est possible dans la mesure où le modèle appris contient de nombreux indices de focalisation. Une recherche exhaustive et non pertinente de toutes les parties à la fois et sans *a priori* est ainsi limitée.

La manière dont l'algorithme effectue la recherche des parties, afin d'être mieux guidé vers l'objet, est complètement définie par la stratégie de reconnaissance choisie. La stratégie est celle du système de perception attentionelle proposé dans ce mémoire. Le processus récursif, de type *génération vérification d'hypothèses*, est guidé par le modèle probabiliste de l'objet. A chaque validation, le modèle est mis à jour ce qui réduit l'espace de recherche et contraint ainsi les détections futures.

La figure 6.2 présente le schéma simplifié qui décrit, de façon générale, la stratégie développée pour le contrôle du processus de reconnaissance. Cette stratégie est constituée de plusieurs étapes qui vont, de la sélection de la partie à reconnaître, jusqu'à la mise à jour du modèle (focalisation) ; en passant par d'autres étapes intermédiaires comme celle de la détection.

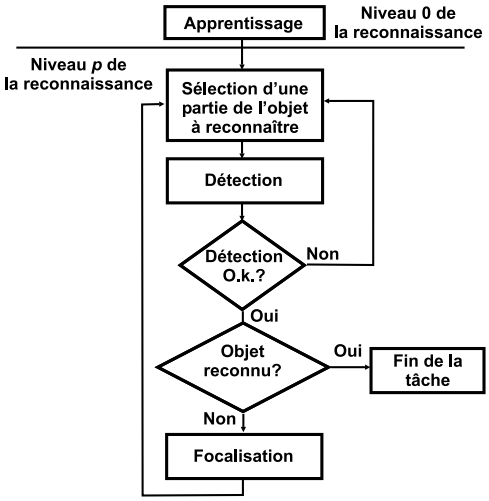


FIG. 6.2 – Organigramme simplifié de la stratégie du système de reconnaissance par focalisation progressive.

Préparation et initialisation : niveau zéro

Une fois l'apprentissage fait, le résultat obtenu correspond au modèle moyen initial de l'objet $\mathcal{N}(\bar{\mathbf{x}}, C_{\mathbf{x}})_0$, dont le sous-indice 0 est donné pour le niveau zéro du processus. Le modèle initial définit pour chaque partie la valeur moyenne de ses paramètres (avec notamment sa position moyenne dans l'image) et leur dispersion. Ceci permet de définir la zone d'intérêt de chaque partie dans l'espace des caractéristiques (y compris donc dans le plan image).

Génération des hypothèses (sélection des *parties*)

Effectuer la génération d'hypothèses revient à hiérarchiser les *parties* composant l'objet pour le sélectionner.

Ici, la partie la plus pertinente est celle qui a le plus d'indices d'attention visuelle, la variance moindre, l'indice de corrélation maximum et le temps de calcul probable le plus faible (privilégier les *parties* de résolution la plus basse).

La procédure est la suivante : à l'état k du processus, la meilleure partie selon ce critère est recherchée parmi toutes celles qui ne sont pas encore sélectionnées. En fait, elle correspond à la partie la plus pertinente dans l'état courant k du processus.

Phase de détection des *parties*

Zone d'intérêt Avant de réaliser une détection, il faut d'abord définir la région d'intérêt (**ROI**) dans l'image où la partie se situe le plus probablement. Cette **ROI** dans l'espace image peut être obtenue à partir de la valeur moyenne sur la position de chaque partie Λ_m, $\overline{\mathbf{a}}_m$ et de sa matrice de covariance $\mathsf{C}_{\mathbf{a}_m}$.

Détection La détection d'une partie Λ_m prend tout d'abord en compte le type de paramètres qui la composent. Puis, la **ROI** définie précédemment est traitée par le sous ensemble des N_m opérateurs.

Décider si une partie est reconnue ou non dans la zone d'intérêt revient à définir des bornes sur la fonction de distribution. Cependant, pour une partie Λ_m, de vecteur de paramètres moyen $\overline{\lambda_m}$ et de covariance C_{λ_m}, il y a L_m vecteurs candidats $\hat{\lambda}_l$; $l \in [1, L_m]$. Un candidat $\hat{\lambda}_l$ est retenu si $(\overline{\lambda_m} - \hat{\lambda}_l)^t \mathsf{C}_{\lambda_m}^{-1} (\overline{\lambda_m} - \hat{\lambda}_l) \leq s$.

Cette relation (distance de Mahalanobis) détermine si le vecteur $\hat{\lambda}_l$ est ou n'est pas à l'intérieur de l'ellipsoïde d'incertitude de la partie dans l'espace des caractéristiques. L'indice $l \in [1, L_m]$ indique le nombre de candidats possibles pouvant correspondre à λ_m, où L_m est le nombre total de candidats pour Λ_m. La valeur de s est fixé à deux (compromis entre risque de non détection et ambiguïté de détection).

Dans le cas où il y a au moins une détection parmi tous les candidats $\hat{\lambda}_l$, le détecteur retourne celle qui est la plus proche de la valeur moyenne $\overline{\lambda_m}$. Si la partie détectée ne correspond pas à la partie de l'objet en question, un autre candidat est évalué, et ainsi successivement si besoin jusqu'au test total des candidats.

Focalisation dans l'espace des caractéristiques : mise à jour du modèle par filtrage de Kalman

Dans le cas général, la région d'intérêt, qui n'est pas à une simple zone géométrique mais une sous partie de l'ensemble de l'espace des paramètre de λ_m est déduite de la matrice de covariance $\mathsf{C}_{\mathbf{x}}$ (voir §6.2.5). Le lien statistique entre ses différents paramètres étant exprimé dans cette matrice de covariance, il est possible, à l'aide d'une simple détection d'une des

parties λ_m, de réajuster non seulement la position des autres parties mais aussi de préciser l'intervalle de variation des paramètres de celles-ci en fonction de cette détection.

En effet, en considérant l'observation $\hat{\lambda}_m = [\hat{\zeta}_{m1}, \ldots, \hat{\zeta}_{mn}, \hat{\mathbf{a}}_m]^t$ de $\lambda_m = [\zeta_{m1}, \ldots, \zeta_{mn}, \mathbf{a}_m]$, pour $n \in [1, N_m]$, il est possible d'écrire les équations d'état suivantes :

$$\underline{x}_n(k+1) = \mathbf{A}_n \underline{x}_n(k) + \mathbf{B}_n \underline{U}_n(k) + \underline{v}_n(k) \tag{6.1}$$
$$\hat{\zeta}_{mn}(k) = \mathbf{H}_n \underline{x}_n(k) + \underline{w}_n(k) \tag{6.2}$$

et

$$\underline{x}_{\mathbf{a}}(k+1) = \mathbf{A}_{\mathbf{a}} \underline{x}_{\mathbf{a}}(k) + \mathbf{B}_{\mathbf{a}} \underline{U}_{\mathbf{a}}(k) + \underline{v}_{\mathbf{a}}(k) \tag{6.3}$$
$$\hat{\mathbf{a}}_m(k) = \mathbf{H}_{\mathbf{a}} \underline{x}_{\mathbf{a}}(k) + \underline{w}(k)_{\mathbf{a}} \tag{6.4}$$

pour les N_m paramètres supposés indépendants issus des opérateurs et les coordonnées correspondantes à la partie Λ_m, respectivement.

Ici, les équations 6.1 et 6.3 (équations d'évolution) n'ont pas d'intérêt : le vecteur $\bar{\mathbf{x}}$ n'évolue pas d'une détection à l'autre, donc $\mathbf{A} = \mathbf{I}$ (matrice identité). Il serait néanmoins possible, dans le cadre d'un processus dynamique prendre en compte l'évolution de la position des *parties* (relation 6.3). Ce ne n'est pas le cas ici. De même la matrice de covariance **Q** du vecteur de bruit d'évolution v est nulle.

S' il y a eu une détection, il est nécessaire de mettre à jour le modèle $\mathcal{N}(\bar{\mathbf{x}}, \mathrm{C}_{\mathbf{x}})_k$ (où k indique l'état courant du processus de reconnaissance) qui s'affine ainsi et permet de mieux cerner la recherche des *parties* futures. Cette mise à jour est effectuée N_m fois en utilisant un filtre de Kalman.

La matrice $\mathbf{H}_i$ est telle que $\hat{\mathbf{x}}_i = \mathbf{H}_i \mathbf{x}$. Elle représente le lien entre $\hat{\mathbf{x}}_i$ et la totalité du vecteur $\mathbf{x}$. $\hat{\mathbf{x}}_i$ est l' estimée du vecteur de paramètres de la caractéristique i, et $\mathrm{C}_{\hat{\mathbf{x}}_i}$ la matrice de covariance des paramètres de la caractéristique i.

Exemple d'évolution

Un exemple d'évolution du processus de reconnaissance est présenté figure 6.3. Il enchaîne 4 itérations du processus. Ces itérations correspondent au cas où l'algorithme a trouvé la bonne hypothèse.

La figure 6.3-a montre la position du modèle moyen initial obtenu lors de l'apprentissage. Dans la première colonne (figure 6.3-b), l'ellipse correspond à la région d'intérêt sur la position probable de la partie en question. Dans la deuxième colonne (figure 6.3-c) est affichée la position de la détection la plus proche du vecteur moyen à l'intérieur de l'ellipse. Dans la troisième colonne (figure 6.3-d) est affiché le modèle mis à jour après la détection des *parties*. Il apparaît que la région d'intérêt des *parties* manquantes est réduite après chaque mis à jour du modèle (figure 6.3-d), ainsi que l'adaptation du modèle à l'objet (voir figure 6.3-ROI et la mise à jour). La figure fait apparaître uniquement la réduction de l'aspect géométrique de la ROI. Par application de ce processus focalisant, c'est la ROI dans la totalité de l'espace des paramètres (incluant les paramètres géométriques) , qui est réduite.

Le fait de resserrer de plus en plus le modèle, entraîne une augmentation du rapport signal/bruit pour les détections futures car le nombre de candidats possibles, pour une partie déterminée, est réduit significativement avec la diminution de l'espace de recherche.

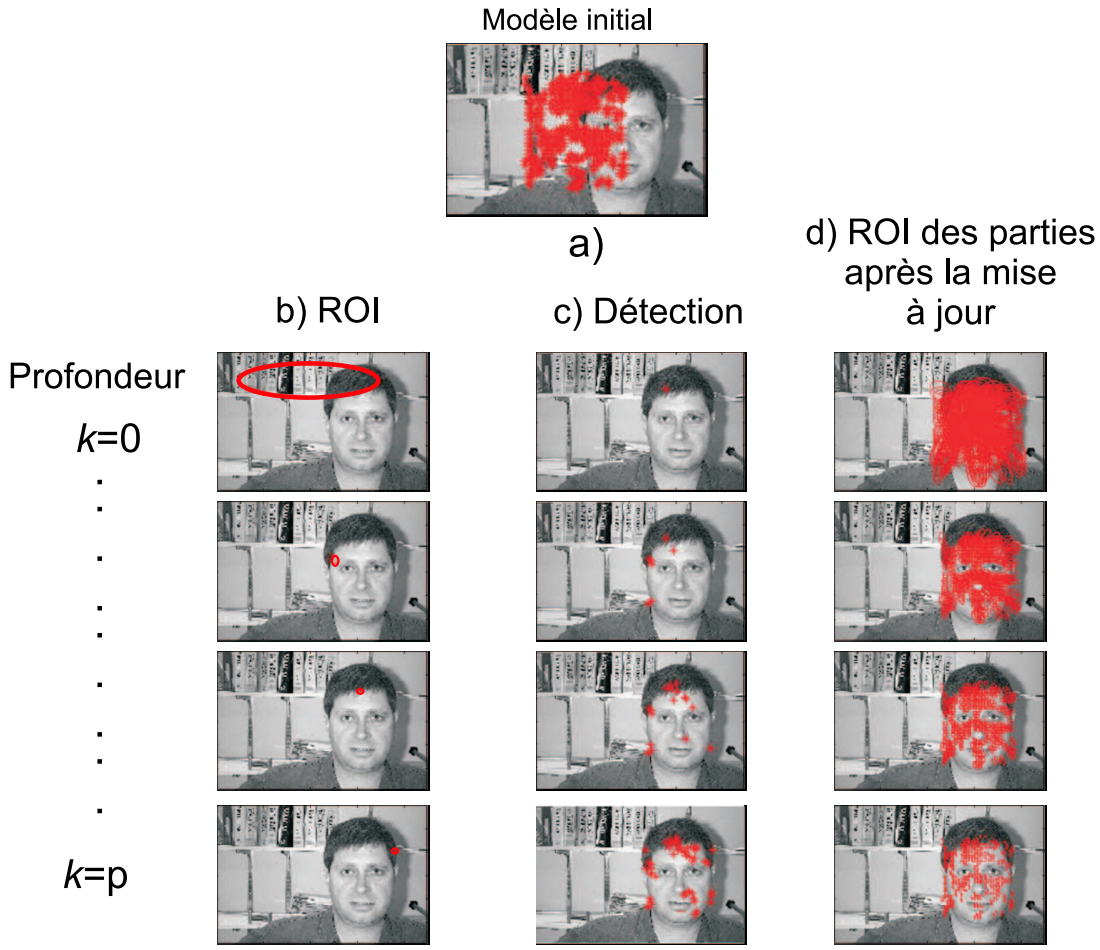


FIG. 6.3 – Exemple d'évolution de la reconnaissance

6.2.6 Phase de décision

Caractérisation probabiliste de la vraisemblance de l'hypothèse

L'étape suivante consiste à savoir à partir de combien de *parties* détectées l'objet est considéré comme reconnu. Ceci revient à déterminer un indice de la pertinence à continuer de chercher d'autres *parties*, selon l'état du processus de reconnaissance. Comme indiqué dans la phase d'apprentissage, chaque partie est caractérisée par un poids indiquant l'importance de cette partie par rapport à l'objet. Il est donc bien évident que le critère de décision doit dépendre de cet indice d'importance mais pas uniquement : la possibilité de détecter ou non une partie tout au long du processus est importante. En effet, le fait de ne pas avoir détecté une partie donnée peut dévalider complètement l'hypothèse courante.

Le processus d'évaluation de la vraisemblance de l'hypothèse courante est modélisée en termes de probabilité conditionnelle.

Le problème peut être posé de la façon suivante : soit une zone d'analyse dans laquelle une partie d'un objet est recherchée et soient :

– d : l'événement "une détection compatible avec la partie recherchée est réalisée",
– O : l'événement "la partie est présente dans la zone d'analyse".

Il est important de considérer non seulement $\Pr(O|d)$ mais aussi $\Pr(O|\overline{d})$.

Alors, $\Pr(O|d)$ (*probabilité a posteriori*) ce qui signifie *la probabilité que l'objet soit présent dans la zone d'analyse sachant qu'on a une détection compatible avec la partie recherchée* :

$\Pr(O|d) = \frac{\Pr(d|O)\Pr(O)}{\Pr(d)}$.

Après une détection donnée, une mesure typique pour savoir si l'objet est présent consiste à évaluer le logarithme du rapport de la probabilité (le rapport de vraisemblance) que l'objet soit présent pour cette détection, sur la probabilité qu'il ne le soit pas.

$$\mathcal{L} = \log\left(\frac{\Pr(O|d)}{\Pr(\overline{O}|d)}\right) = \log\left(\frac{\Pr(d|O)}{\Pr(d|\overline{O})}\right) + \log\left(\frac{\Pr(O)}{\Pr(\overline{O})}\right) \tag{6.5}$$

Avec :
– $\Pr(O)$: la probabilité que l'objet se trouve en effet dans la zone d'analyse. Pour la première itération, une valeur fixe (par exemple 0,5) est choisie a priori. Ensuite, la zone est réduite au fur et à mesure de la détection des *parties*, ce qui entraîne une augmentation de $\Pr(O)$.
– $\Pr(\overline{O})$: elle est simplement donnée par $1 - \Pr(O)$.

Le calcul de la probabilité d'avoir une détection de la partie recherchée dans les deux cas possibles, ($\Pr(d|O)$ et $\Pr(d|\overline{O})$), est décrite dans la suite.

probabilité a priori $\Pr(d|O)$ **.** Il s'agit de la probabilité qu'a le détecteur de trouver une partie compatible avec Λ_m, dans la zone sachant que l'objet s'y trouve. Pour cette valeur il faut prendre en compte de manière plus précise les choses : dans le cas où la partie est présente, la probabilité que le détecteur réponde peut être estimée en prenant en compte les données d'apprentissage.

Branch & Bound

Si la probabilité de présence de l'objet devient très faible, il convient d'arrêter le parcours de la branche correspondante de l'arbre des hypothèses. Le critère d'arrêt est donné par : $\mathcal{L}_k \leq \tau_s$, où τ_s est un seuil défini par l'utilisateur.

Si ce test est vrai, le processus élimine l'hypothèse courante et revient en arrière dans l'arbre de recherche pour en tester une nouvelle.

Décision finale par SVM

Après certaines détections, les régions d'intérêt, dans l'espace géométrique, sont très petites pour les parties restantes. Dans ce cas, il peut être plus pertinent d'arrêter la récursivité et de prendre une décision dans l'ensemble de l'objet. Un classificateur de type SVM est alors utilisé pour prendre la décision finale : objet/non objet.

Suivi

Le modèle de l'objet sur lequel s'appuie le processus attentionnel est dans le cas général issu de la phase d'apprentissage. Dans le cas de l'analyse d'une séquence temporelle d'images, il est possible de piloter le processus attentionnel sur une image, non pas avec le modèle moyen issu de l'apprentissage mais à partir du modèle résultant de la dernière mise à jour de l'image précédente.

La reconnaissance se focalise alors non pas sur une instance générique de l'objet (excepté sur la première image) mais sur l'instance particulière obtenue à la fin du traitement sur l'image précédente.

La méthode apporte donc une solution au problème du suivi d'objet sans aucune modifications de sa structure.

6.3 Résultats

6.3.1 Résultat sur des images seules

Afin de valider l'approche, plusieurs tests ont été faits dans le cadre de la reconnaissance de visages (fig. 6.4) et de piétons (fig. 6.5). Le modèle est placé correctement sur l'objet, en position et en échelle.

Le tableau 6.1, résume la statistique sur le nombre maximal des évaluations SVM et des mises à jour (MAJ), pour le test de reconnaissance de visages. Pour ce test, trois modèles différents ont été utilisés : le modèle tel qu'issu de la phase d'apprentissage (modèle 1), et deux modèles réduits en prenant les 150 (modèle 2) et 50 (modèle 3) *parties* les plus corrélées.

	Modèle 1	**Modèle 2**	**Modèle 3**
nb. de *parties*	300	150	50
nb. de MAJ	46,25	38,28	18
nb. évaluations SVM	9	6	4

TAB. 6.1 – Moyenne du nombre de mises à jour du modèle et d'évaluations SVM, pour trois modèles différents.

Bien que, pour les trois tests, il n'y ait pas une variation significative en nombre de MAJ et d'évaluations SVM, il y a une réduction notable du temps de calcul (un facteur de 60 entre le modèle 1 et le modèle 3).

6.3.2 Suivi d'objet

La figure 6.6 illustre la capacité naturelle de la méthode à suivre un objet particulier (ici un piéton en environnement intérieur) . Elle illustre aussi son comportement face au changement d'échelle de l'objet. Le modèle s'adapte tout aussi bien à l'objet s'il est proche de la caméra que s'il en est très éloigné.

FIG. 6.4 – Exemple de reconnaissance de visages.

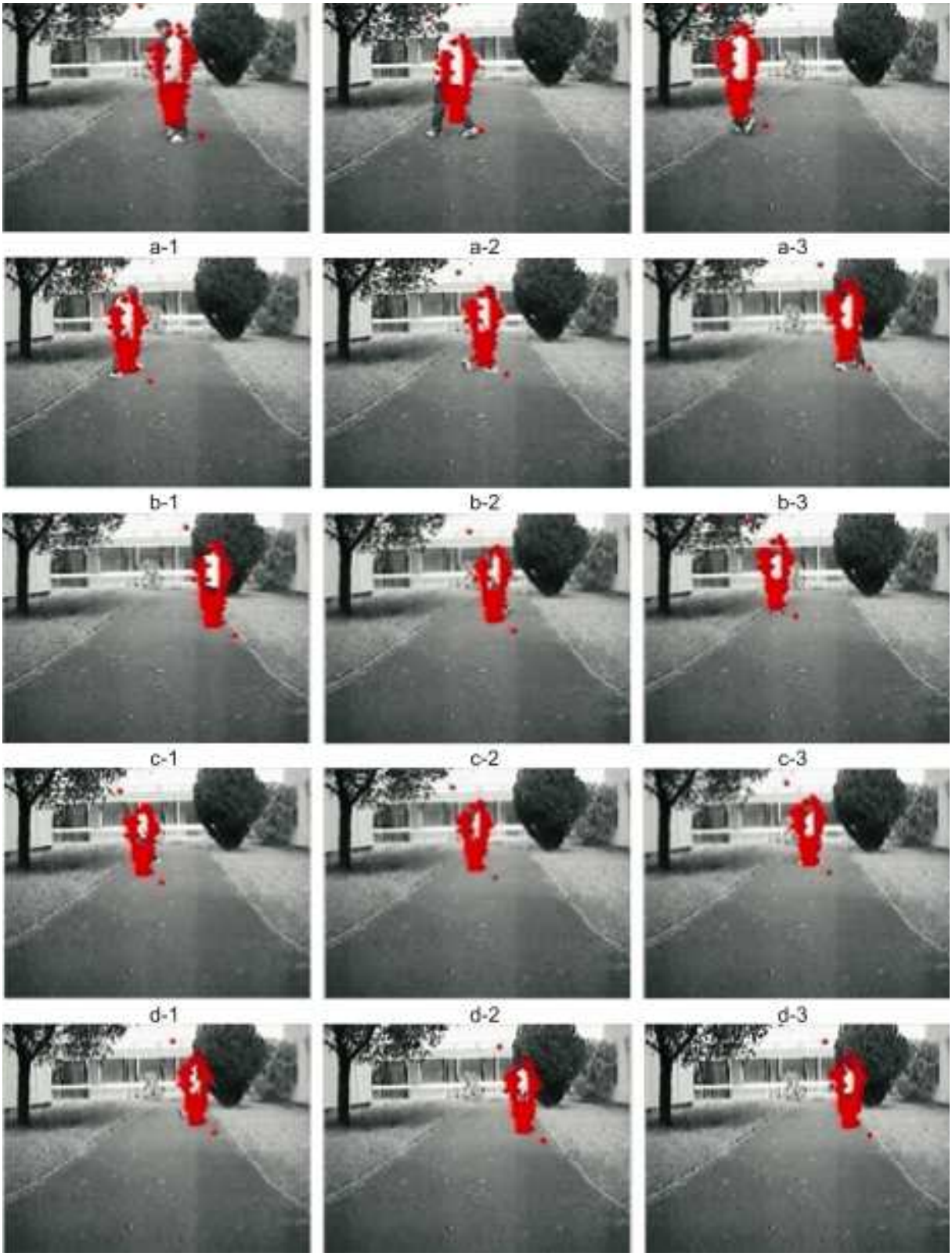


FIG. 6.5 – Exemple de reconnaissance de piétons.

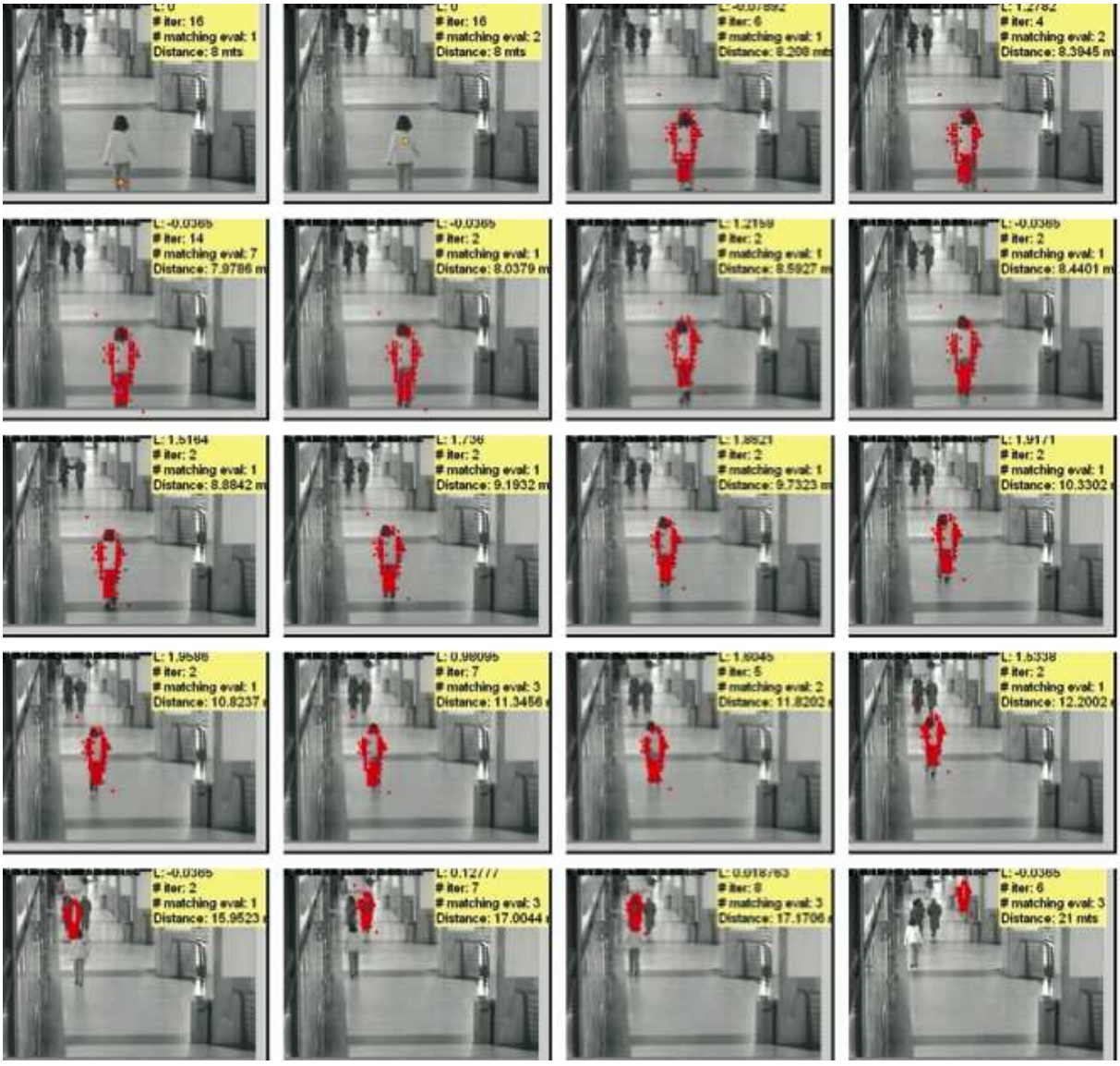


FIG. 6.6 – Exemple de suivi de piéton.

6.4 Bilan

La mise en œuvre de la méthode de perception attentionnelle progresse vers l'analyse globale de scène. La reconnaissance d'objet par vision focalisée est une étape importante vers ce but. Cette approche constitue une modélisation statistique et probabiliste de l'attention visuelle. Elle se veut sinon générique du moins largement généralisable. Les résultats obtenus sur des classes d'objets différentes viennent à l'appui de cette remarque.

La tâche de reconnaissance est considérée comme un processus de perception actif. Le processus attentionnel et la reconnaissance sont contrôlés par un même processus ce qui constitue une originalité importante de ce travail. Le contrôle s'appuie sur un modèle appris de l'objet. Cet apprentissage fait intégralement partie de la méthode. L'apprentissage détermine les caractéristiques les plus pertinentes et construit ainsi le modèle de l'objet. Le modèle se construit en incluant des caractéristiques structurelles comme des caractéristiques d'apparence ce qui permet de tirer le meilleur parti des avantages de chacun de ces deux types de caractéristique.

Le processus attentionnel de focalisation dans l'espace des caractéristiques mis en avant par Tsotsos est intégré dans l'approche : sélection des opérateurs détecteurs, de l'intervalle possible de réponse des détecteurs, de la résolution et sélection de la zone géométrique d'analyse de données du capteur. Cette focalisation permet d'éviter certains écueils des approches classiques en évitant une recherche exhaustive en position et en échelle de l'objet dans l'image.

Enfin il faut remarquer que cette approche intègre la détection, la reconnaissance et le suivi d'objet dans un formalisme unique.

6.5 Perception multisensorielle des piétons

6.5.1 Introduction

En 2005, plus de 7000 piétons ont été tués dans des accidents de la circulation au sein de l'union européenne. Les statistiques des accidents indiquent qu'en dépit des récents progrès réalisés en matière de sécurité, grâce à l'instauration des systèmes de sécurité passifs et au renforcement des lois pour les piétons, ces derniers représentent encore la deuxième source la plus importante de blessés et de tués de la circulation routière, après les accidents impliquant les passagers de véhicules [39]. Actuellement, la recherche de l'industrie automobile porte une attention particulière aux systèmes de sécurité actifs qui permettraient de réduire le nombre et la gravité des accidents de piétons.

La détection et la classification des piétons sont des processus complexes. Les données issues des capteurs embarqués sont analysées pour rechercher les caractéristiques qui indiquent la présence d'entités pouvant être des piétons au sein d'une zone d'observation. Ces caractéristiques sont ensuite analysées à l'aide de différentes techniques afin de déterminer si elles représentent ou non des piétons, malgré les différentes variantes de formes et de tailles de piétons.

Le capteur naturelement utilisé pour effectuer la détection est la caméra vidéo. Les systèmes basés sur la vision sont couramment utilisés pour la détection des piétons [89] et [14], car les caméras offrent un niveau de résolution suffisamment élevé pour produire une classification fiable. En outre, le coût des caméras et leur utilisation potentielle pour d'autres fonctions de perception dans le véhicule leur confèrent un avantage supplémentaire par rapport à d'autres types de capteurs. En revanche, l'inconvénient des systèmes de détection basés uniquement sur l'analyse d'image réside moins dans le coût de calcul élevé lié à la classification d'un grand nombre de régions candidates que dans le manque de fiabilité. De plus, dans les environnements extérieurs, la performance des systèmes basés sur la vision est affectée de manière négative par les conditions climatiques. C'est la raison pour laquelle un capteur supplémentaire peut avantageusement être associé à la caméra [38]. L'évaluation de l'état de l'art en matière de détection des piétons par les capteurs effectuée par Gravilla [39] a démontré que les télémètres laser combinés aux caméras représentent une solution avantageuse à exploiter. Ces deux capteurs ont fait la preuve de leur capacité à offrir de bonnes performances dans les systèmes de détection des piétons lorsqu'ils sont utilisés séparément [8] ,[66] et [37], et de performances encore meilleures lorsqu'ils sont utilisés ensembles [76] et [57].

Dans certaines solutions, [76] seule la complémentarité des deux capteurs est exploitée. Le système de détection des piétons fait alors appel à un télémètre laser pour localiser les objets et calculer les régions candidates. Un classifieur d'images sur la base d'un réseau de neurones convolutif est ensuite appliqué pour classer ces candidats. Cette association leur permet de développer un système de détection des piétons en temps réel puisque la durée nécessaire à la segmentation de l'image est considérablement réduite. Les inconvénients sont que les piétons non détectés par le télémètre laser ne peuvent pas être détectés par la vision ; et que les performances du système sont fortement liées à la fiabilité du classifieur basé sur la vision. Ce dernier point a été amélioré [57]. Les méthodes de classification d'objets fonctionnent à la fois dans l'espace laser (vote majoritaire associé à un classifier à mixture de gaussienne) et dans l'espace de vision (classifieur Adaboost). Une règle de décision basée sur la règle de Bayes est utilisée afin de combiner les résultats de chaque technique de classification et d'obtenir ainsi un objet de classification plus fiable.

L'objectif consiste à concevoir un système, à l'usage des automobiles grand public, capable de détecter et de pister le passage de piétons devant un véhicule. La solution proposée doit être rapide, simple à mettre en œuvre et fiable. L'architecture du système de perception est représentée 6.7. Il repose sur l'association d'une caméra et d'un télémètre laser. La détection et le pistage sont effectués dans l'espace laser uniquement. Les deux capteurs sont calibrés. La classification s'effectue d'abord dans les données du télémètre puis dans l'image. La classification dans l'espace du télémètre laser repose sur des règles heuristiques concernant des profils de piétons et sur leurs contraintes cinématiques. En complément, le classifieur basé sur la vision utilise l'algorithme Adaboost avec des caractéristiques du type ondelettes de Haar décrites dans [67]. Les décisions des classifieurs sont combinées à l'aide de la règle de Bayes. Les probabilités sont estimées à partir des scores en sortie des classifieurs de chaque capteur. Le système s'adapte à la taille de l'objet détecté. La variation est incluse de manière originale dans dans la probabilité de la vision. Ceci permet de réaliser une meilleure classification des piétons petits et éloignés. Enfin, l'intégration des connaissances passées augmente la fiabilité du système de perception des piétons, en réduisant le nombre d'erreurs de détection.

Cette application implémente une version simplifiée du système de perception attentionnel par focalisation. Son schéma de principe est représenté figure 6.7 Elle correspond à un contexte multisensoriel et introduit de manière explicite un caractère *bottom-up* dans le système attentionnel général. La représentation synthétique de la scène censée guider la perception est initialement vide. Elle évolue dynamiquement en fonction des obstacles détectés par le télémètre LASER. Une fois ceux-ci positionnés dans la carte, le processus de focalisation guide l'analyse d'image qui dans ce cas n'a pas pour rôle la mise à jour du modèle mais la seule évolution de la vraisemblance de l'hypothèse. A ce sujet, le formalisme bayésien sur lequel repose le calcul de la vraisemblance est mis en œuvre de manière originale dans le cas multisensoriel. Ce travail correspond la thèse de Laurence NGAKO-PANGOP en cours de préparation (soutenance prévue en 2008) que je co-encadre à 33%. Cette thèse est financée par une bourse CIFRE avec le constructeur automobile Renault. Elle est soutenue de plus par un contrat de recherche de 30k€.

6.5.2 Description détaillée de l'application

Détecteurs

Extraction des caractéristiques du télémètre laser Cette section décrit la méthode et le calcul de segmentation des caractéristiques extraites des données du télémètre laser. Les impacts télémétriques sont regroupés en clusters : deux impacts appartiennent au même cluster si la distance qui les sépare est inférieure à un seuil. Bien qu'il ait été possible d'envisager des seuils plus complexes ou adaptifs, les expérimentations montrent que ce simple schéma de segmentation fonctionne parfaitement bien.

Un processus d'extraction des caractéristiques stable est fondamental pour assurer la fiabilité et la robustesse de l'ensemble du système. Sous l'hypothèse que l'approximation polygonale est valide, l'ajustement linéaire des clusters est adapté à l'approximation des profils des objets. Le processus d'ajustement linéaire utilisé est détaillé dans [55]. De plus un objet de type piéton est supposé produire un segment court. Ainsi, le processus d'ajustement linéaire permet de rejeter les objets improbables. La figure 6.8 montre le résultat de ce détecteur avec des objets de type piéton et des objets rejetés. La caractéristique conservée ici est la largeur de l'objet de type piéton. Cette caractéristique est la distance Euclidienne entre le premier et le dernier

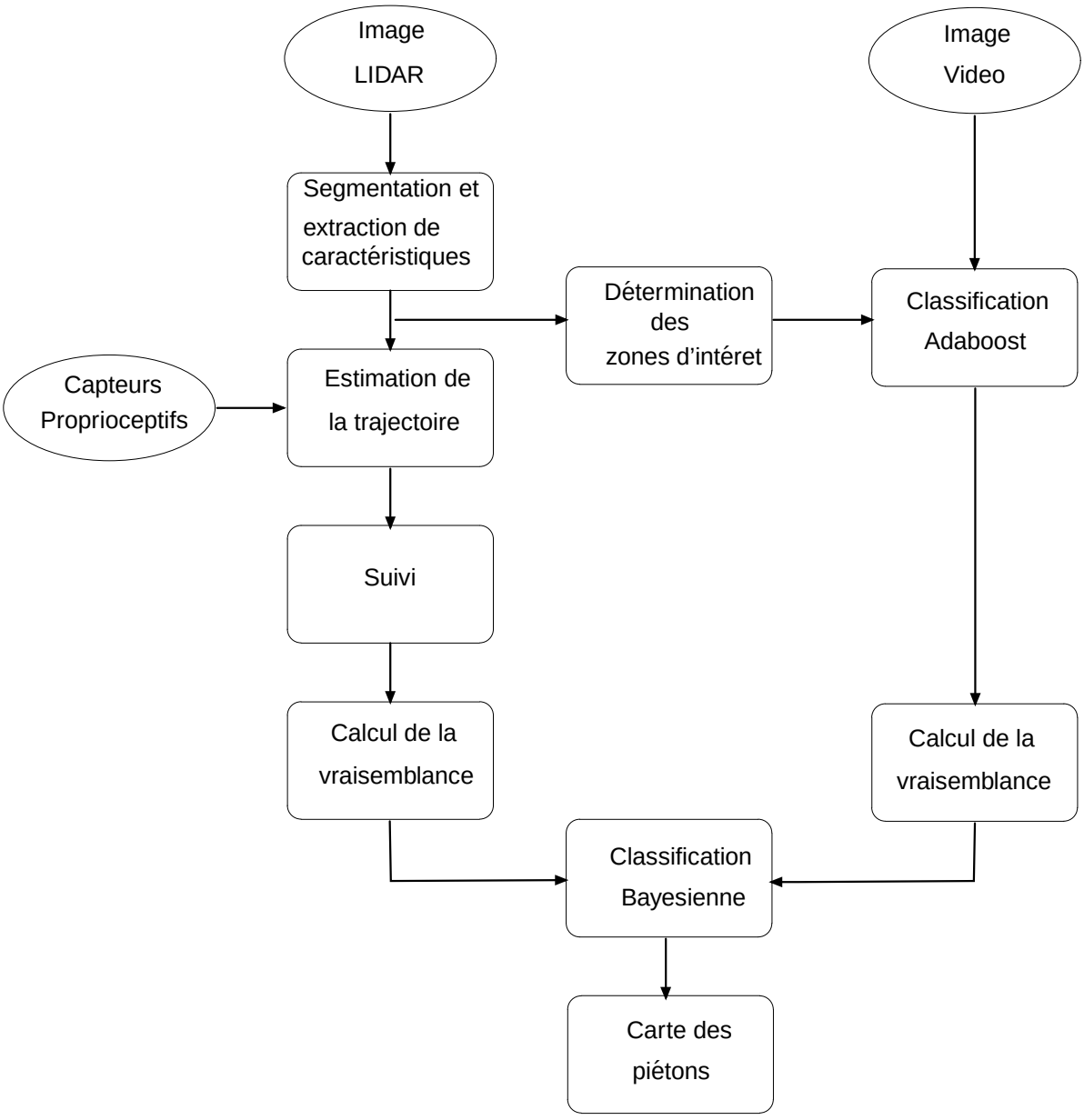


FIG. 6.7 – Système de perception utilisant un LIDAR et des informations vision pour la classification et la détection de piétons.

point d'un segment.

Un filtre de Kalman est utilisé pour effectuer le pistage de l'objet dans un repère lié au LIDAR qui pour cette application constitue le repère absolu. Le modèle de mouvement du piéton est présenté avec une vitesse constante et une accélération à marche aléatoire. Le centroïde de chaque cluster est calculé et utilisé en tant que point caractéristique, c'est-à-dire que le comportement de l'objet est décrit par rapport à son centroïde. L'algorithme de pistage effectue une estimation de l'état (la position et les composantes du vecteur vitesse) des objets à partir de l'observation actuelle et une prévision de l'état (à partir de la précédente estimation). La vitesse du piéton est également estimée par le filtre de Kalman car elle peut constituer une mesure discriminante en matière de classification des objets.

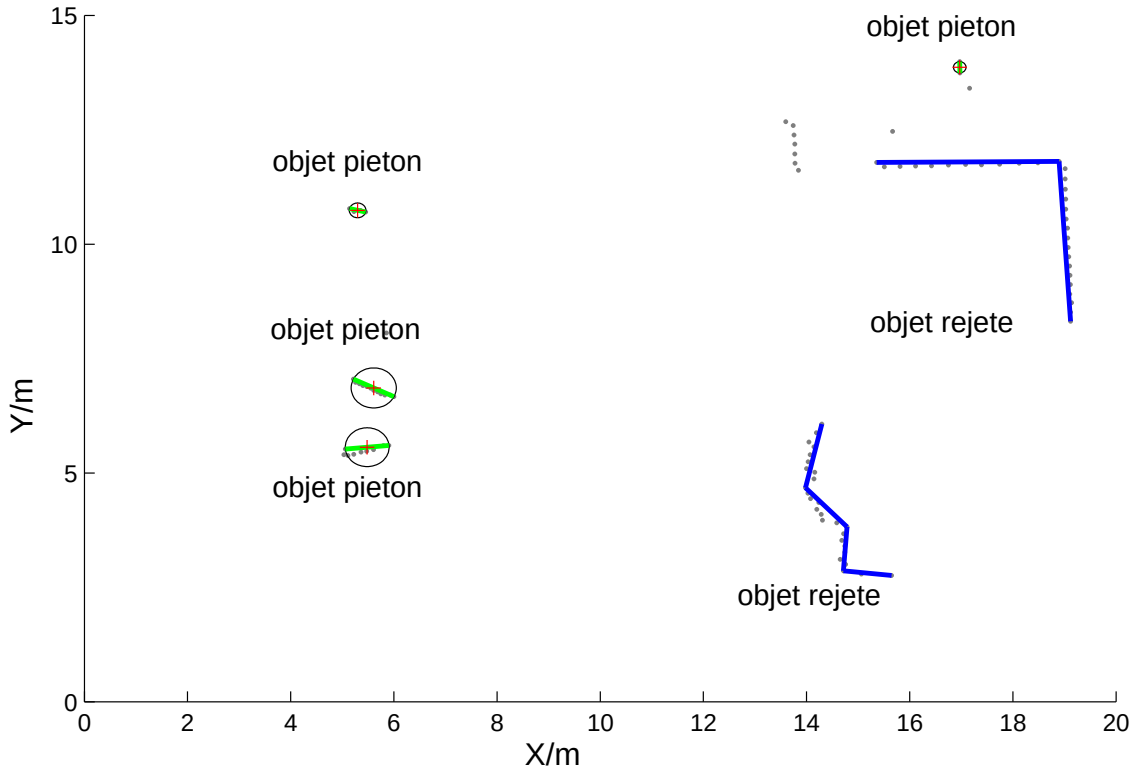


FIG. 6.8 – Illustration de la procédure d'ajustement linéaire : des considérations liées au contour permettent une classification grossière.

Caractéristique vision : sortie du classifieur Adaboost Le classifieur Adaboost[2] est un algorithme d'apprentissage statistique [35]. Le modèle de piéton est obtenu par l'analyse de deux jeux d'images d'apprentissage. Le premier jeu est composé de plusieurs instances de piétons, appelées échantillons « positifs ». En revanche, le deuxième jeu d'images ne contient pas de piétons, il est constitué d'échantillons « négatifs ». Plusieurs caractéristiques sont extraites de chaque échantillon d'apprentissage. L'objectif principal de l'algorithme Adaboost consiste à rechercher un très petit nombre de ces caractéristiques pouvant être combinées pour former un classifieur efficace. Pour le mettre en pratique, un algorithme d'apprentissage faible est conçu pour sélectionner l'unique caractéristique qui distingue le mieux les échantillons positifs des échantillons négatifs. Le détail de la mise en œuvre de de l'algorithme Adaboost pour la reconnaissance d'objet est précisé dans [67]. Dans le système décrit ici, le classifieur Adaboost n'est employé que sur l'image et sa sortie est utilisée en tant que caractéristique vision. La détection dans les données LIDAR est tout d'abord effectuée. Son résultat est combiné ensuite à la caractéristique vision pour réaliser l classification finals par le classifieur Bayésien.

Estimation de la vraisemblance : classifieur bayésien

L'objectif du classifieur Bayésien est de calculer pour chaque objet détecté la probabilité que celui-ci a d'être un piéton. Ce calcul est effectué à partir des probabilités de largeur, de vitesse

[2] Adaptative Boosting

et du score Adaboost. La probabilité *a priori* de chaque classe (classe piéton ou non piéton) est initialisée à $0,5$ et varie lorsque l'objet est pisté. Cette section présente également l'intégration de la variation d'échelle dans la probabilité (score) de la vision.

Calcul de la probabilité de présence d'un piéton Pour un vecteur de caractéristiques donné, $X_k = (x_1, \cdots, x_n)^T$, la probabilité a posteriori $P_k(\mathrm{p}|X_k)$ est calculée en considérant le x_i ($i = 1..n$) à l'aide de l'équation suivante :

$$P_k(\mathrm{p}|X_k) = \frac{1}{\beta} p(x_1|\mathrm{p}) \cdots p(x_n|\mathrm{p}) P_k(\mathrm{p}) \tag{6.6}$$

où $P_k(\mathrm{p})$ est la probabilité a priori de la classe piéton à la $k$ième itération, et β une constante de normalisation.

Dans cette application, $X_k = (largeur, vitesse, score)^T$. Pour $C \in \{\text{pieton}, \text{non-pieton}\}$, qui représente les classes piéton et non piéton, il convient de modéliser $p(\text{largeur}|C)$, $p(\text{vitesse}|C)$ et $p(\text{score}|C)$, les fonctions de densité de probabilité conditionnelles à chaque classe.

$p(\text{score}|C)$ est estimée à partir du processus d'apprentissage de l'algorithme Adaboost. Le score suit une distribution normale ($\mathcal{N}$) qui correspond à l'histogramme des scores obtenus lors de la caractérisation de l'algorithme Adaboost.

$$Score \sim \mathcal{N}(\mu, \sigma^2) \tag{6.7}$$

où μ et σ sont respectivement la moyenne et l'écart-type de cette distribution normale.

Les distributions suivantes sont proposées pour $p(\text{largeur}|C)$ et $p(\text{vitesse}|C)$.

La largeur d'un piéton suit une distribution normale paramétrée à partir de l'observation de certaines séquences de piétons en mouvement.

$$Largeur \sim \mathcal{N}(\mu_w, \sigma_w^2) \tag{6.8}$$

La largeur d'un objet non piéton suit une distribution uniforme puisqu'aucune hypothèse *a priori* n'est émise. La largeur minimale, W_{min}, est la résolution du télémètre laser, (ici $5.10^{-2}m$) ; la largeur maximale, W_{max} est la longueur maximale du pas d'un piéton définie égale à $1m$.

$$Largeur \sim \mathcal{U}(W_{min}, W_{max}) \tag{6.9}$$

La vitesse d'un piéton suit une distribution uniforme, car un piéton peut se déplacer à n'importe quelle vitesse dans un intervalle $[V_{min}, V_{max}]$. Ici, $V_{min} = 0m/s$ représente un piéton immobile et $V_{max} = 2m/s$ est la vitesse maximale autorisée pour les piétons concernés.

$$Vitesse \sim \mathcal{U}(V_{min}, V_{max}) \tag{6.10}$$

Les vitesses des objets suivent une distribution tracée sur la figure 6.9. Ce choix est motivé par le fait que la plupart des objets qu'il est possible de confondre avec des piétons sont statiques.

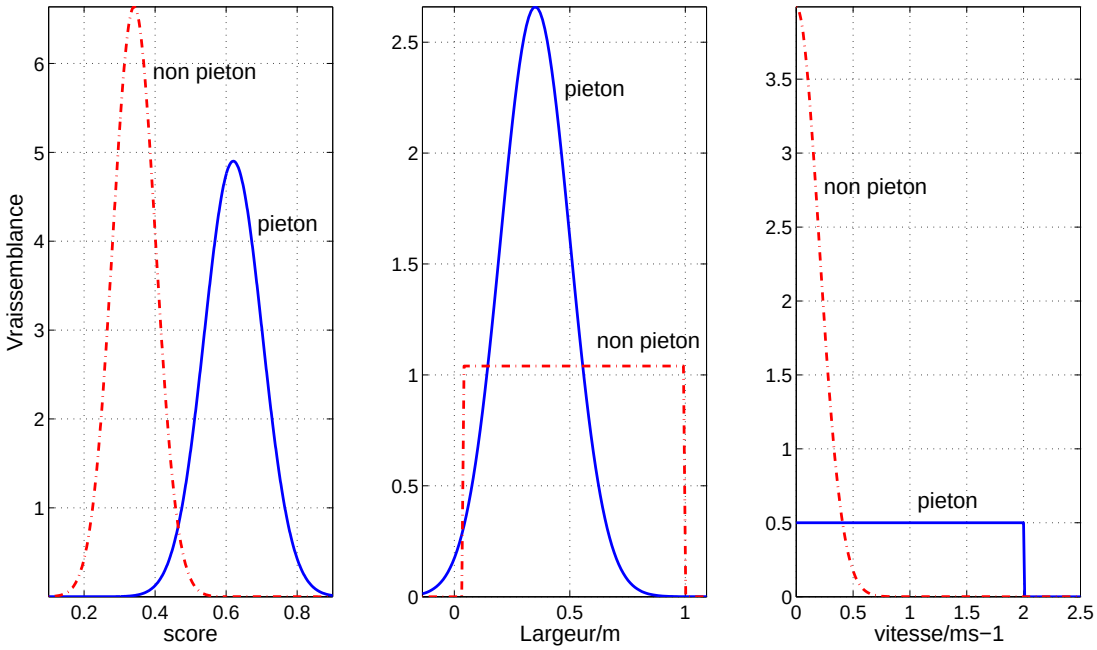


FIG. 6.9 – Densités de probabilité conditionnelles pour les deux classes selon chaque caractéristique.

Influence de la variation d'échelle

Pour détecter des piétons dans une image vidéo, des rectangles englobants sont calculés. Ils peuvent entourer les piétons potentiels aux emplacements donnés par le télémètre laser. La taille de ces rectangles englobants dépend de la distance entre les objets et le véhicule, ainsi que de la hauteur des piétons potentiels. Un redimensionnement est nécessaire avant d'appliquer l'algorithme Adaboost, car le processus d'apprentissage est réalisé avec des échantillons d'images de taille fixe. Lorsque le piéton est éloigné du véhicule ou dans le cas de la détection des enfants (même à proximité du véhicule) les rectangles englobants doivent être petits. Si la taille de l'image du jeu d'apprentissage est supérieure à celle des rectangles englobants, le redimensionnement détériore l'aspect du piéton et le résultat du classifieur basé sur la vision est alors modifié. Par exemple, un même piéton peut avoir un score élevé à proximité du véhicule et un score faible s'il est éloigné. Pour rendre le classifieur Bayésien plus performant face à la variation d'échelle, les paramètres de probabilité de la vision $p(\text{score}|C)$ (moyenne et écart-types) sont adaptés à la taille des cadres englobant. Ainsi, plutôt que de classer automatiquement un piéton lointain ou petit en tant qu'objet non piéton en raison de son score Adaboost faible, le classifieur Bayésien utilise les paramètres de probabilité de la vision qui correspondent à la taille des rectangles englobants et ses résultats sont plus probants. La figure 6.10 montre la correspondance entre le facteur d'échelle (rapport de la largeur du rectangle englobant et de la largeur de l'image du jeu d'apprentissage) et les paramètres de probabilité de la vision.

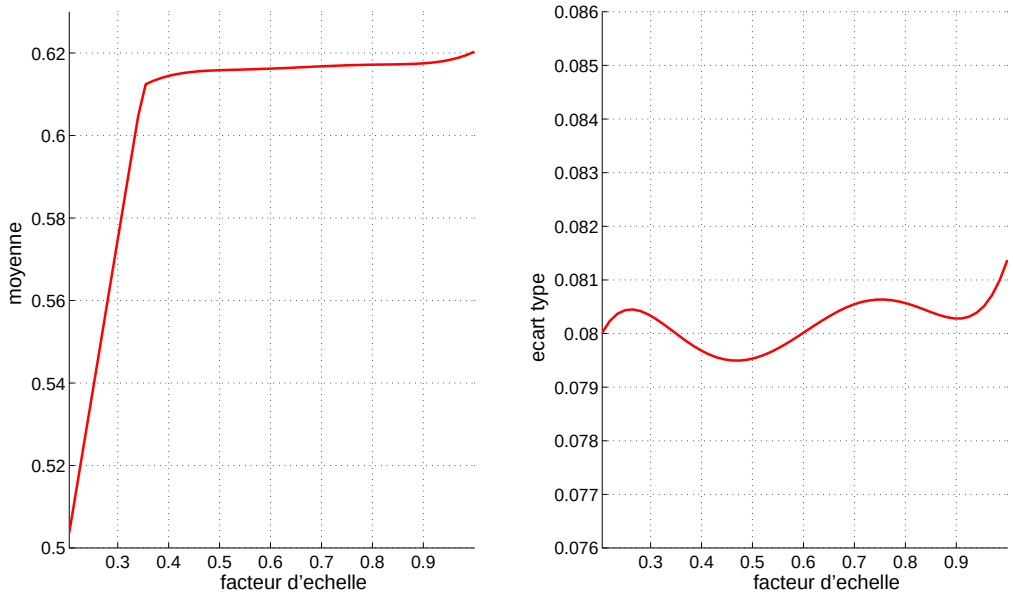


FIG. 6.10 – Paramètres de la densité de probabilité du score de classification Adaboost issu de la vision en fonction du facteur d'échelle.

Intégration des connaissances passées

Contrairement aux méthodes de classification courantes qui procèdent image par image ou utilisent le pistage uniquement pour réduire les erreurs de détection (par exemple, un objet est classé uniquement s'il apparaît dans des images consécutives), la méthode Bayésienne classe un objet dès qu'il est détecté par le télémètre laser. Dans ce cas, la réponse de classification d'un objet pisté inclut les résultats précédents. La relation suivante est utilisée :

$$P_k(\mathrm{p}) = P_{k-1}(\mathrm{p}|X_{k-1}) \tag{6.11}$$

Cette relation signifie qu'une probabilité *a priori* $P_k(\mathrm{p})$ de l'itération courante correspond à la probabilité *a posteriori* $P_{k-1}(\mathrm{p}|X_{k-1})$ calculée à l'itération précédente. Ceci améliore la fiabilité de la classification en permettant effectivement de réduire les erreurs de détection.

6.5.3 Résultats expérimentaux

Dans cette section, sont présentées les expérimentations réalisées devant les bâtiments du LASMEA et sur un passage pour piétons. Le télémètre laser mono-plan utilisé est fixé à l'avant du véhicule à environ $0,4m$ du sol, pour être à hauteur des jambes d'un d'un piéton. La caméra vidéo est une caméra niveaux de gris. Un calibrage des deux capteurs est effectué hors ligne ce qui permet de projeter les objets détectés par le télémètre laser dans l'espace de la caméra. Deux séries d'expérimentations sont présentées sur des séquences de données réelles : la première dans un environnement semi-structuré et la deuxième sur un passage pour piétons, le véhicule est statique dans les deux cas.

Un détecteur Adaboost de base est construit à partir de 100 classifieurs « faibles ». Ce détecteur est préalablement entraîné sur une base de données de 516 images de personnes de face et de dos. Ces personnes sont centrées et approximativement de la même taille, dans des fenêtres de 128 x 64 pixels. En revanche, le jeu d'apprentissage des non piétons comporte 2550 images de scènes naturelles d'une taille de 128 x 64 pixels, sans personne.

Des expérimentations sur des données simulées sont présentées pour comparer les performances du classifieur utilisant la combinaison des caractéristiques (largeur, vitesse et score) avec le classifieur basé sur les informations fournies par la vision ou le laser uniquement. Un grand nombre de triplets de caractéristiques sont générés de manière aléatoire à partir des densités de probabilité conditionnelles de chaque classe. Ces triplets sont ensuite classés comme s'ils étaient mesurés dans un jeu d'apprentissage d'échantillons positifs et négatifs.

Les courbes ROC [3] résultantes sont représentées figure 6.11. Dans l'expérimentation, seules les caractéristiques LASER (largeur et vitesse) sont présentes dans la classification. La courbe en trait continu bleue représente le classifieur basé sur la vision. Pour ce qui concerne les courbes "classifieur image" et "classifieur LIDAR", la caractéristique de la vision semble plus discriminante que les caractéristiques LASER pour le classement des piétons. Ceci provient probablement du fait que le modèle LASER est basé sur une heuristique et que le modèle de vision est obtenu par un processus d'apprentissage. L'expérimentation finale est effectuée en utilisant à la fois des caractéristiques laser et de vision. Pour un pourcentage de faux positif de 1%, le taux de détection réussie est de 75% avec le classifieur laser, de 96% avec le classifieur basé sur la vision, et de 98% avec le classifieur combinant les caractéristiques laser et de vision.

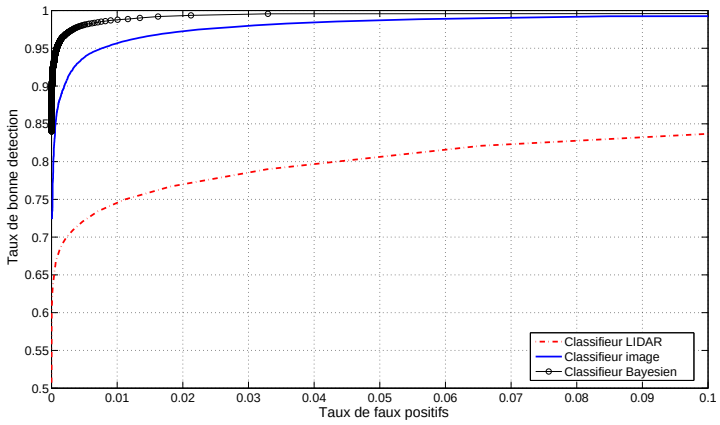


FIG. 6.11 – Courbe ROC optimales. Elles illustrent la pertinence des densités de probabilité de chaque classe.

La première expérimentation sur données réelles est réalisée dans un environnement semi-structuré ; un piéton marche devant un véhicule en stationnement. La figure 6.12 montre une vue de la scène et la trajectoire suivie par le piéton. L'objectif du test consiste à mesurer la pertinence de la méthode de combinaison Bayésienne incluant l'intégration de la connaissance passée et l'influence de la variation d'échelle.

Les caractéristiques extraites dans cette séquence, ainsi que les probabilités de présence des

[3] Receiver Operating Characteristic : courbe de caractéristiques d'efficacité

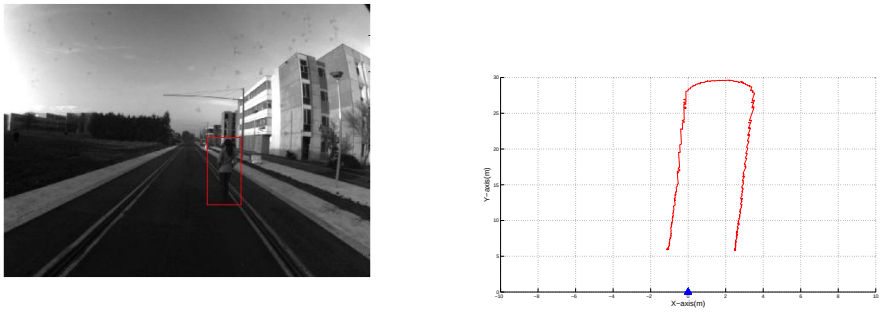


FIG. 6.12 – Vue de la scène et trajectoire du piéton

piétons sont représentées sur la figure 6.13. Les caractéristiques extraites des capteurs sont trop bruitées. Ceci conduit à un échec du classifieur basé sur les capteurs seuls. En revanche, le classifieur Bayésien est plus robuste et plus fiable, car il intègre la connaissance passée, ce qui permet d'harmoniser la probabilité dans le temps.

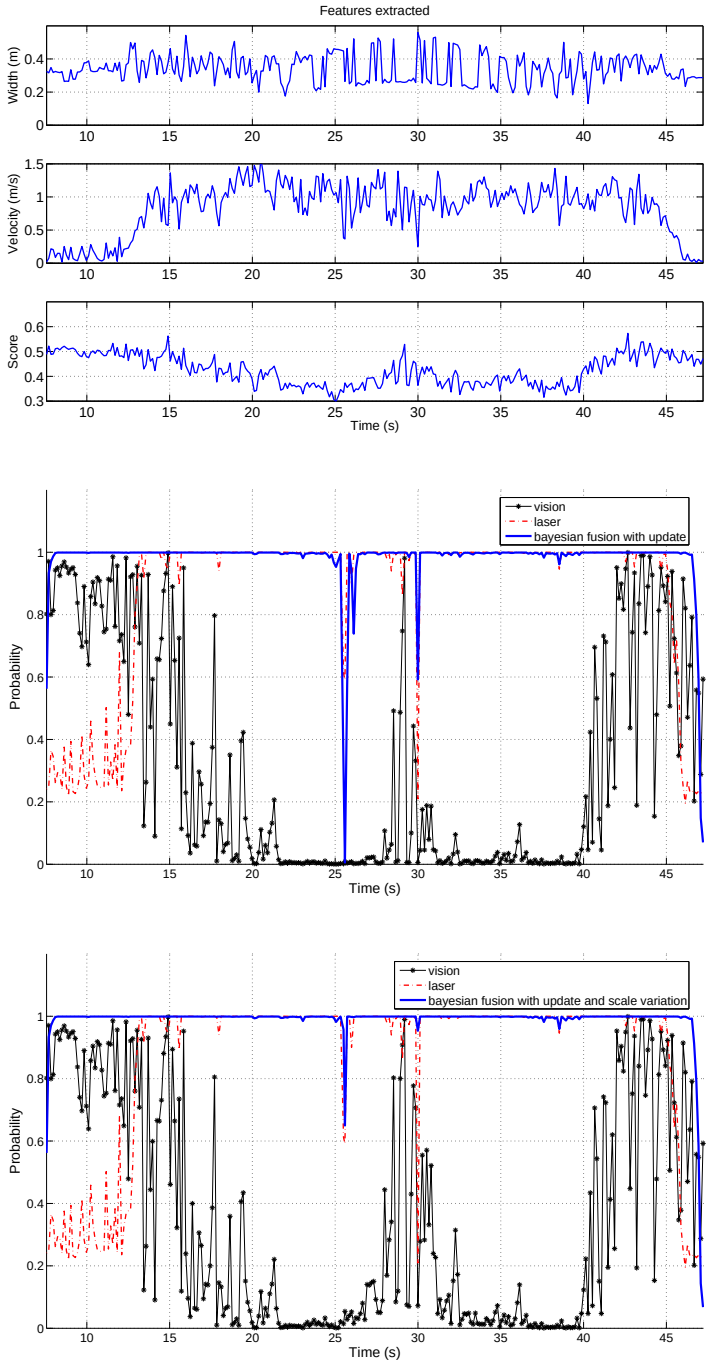


FIG. 6.13 – Evolution temporelle du résultat de la classification des piétons. La prise en compte de la variation d'échelle diminue le nombre de fausses détections.

L'objectif du deuxième test consiste à vérifier le comportement du système dans un environnement encombré. Ici, seuls la largeur et le score sont combinés, car le pistage multi- cible est en cours de développement. La figure 6.14 illustre tous les obstacles détectés par la segmentation du télémètre laser.

FIG. 6.14 – Objets segmentés dans le télémètre LASER.

Les figures 6.15, 6.16 et 6.17 illustrent respectivement les résultats des classifieurs basés sur la largeur, la vision et le classifieur Bayésien. Seuls les objets dont la probabilité d'être un piéton est supérieure à 70% sont mis en évidence.

FIG. 6.15 – Résultat du classifieur basé sur la largeur.

Les erreurs de classification sur la figure 6.17 figurent en vert et les détections réussies figurent en rouge. Dans cette image, le taux de détection réussie est de 100% et le taux d'erreur de détection est de l'ordre de 29%.

FIG. 6.16 – Résultat avec un classifieur adaboost.

FIG. 6.17 – Résultat du classifier Bayesien. Tous les piétons qui traversent sont détectés mais il reste des fausses détections.

6.5.4 Conclusion

Dans cette partie est présentée une méthode Bayésienne pour combiner des caractéristiques extraites d'un télémètre laser et d'une caméra vidéo à des fins de classification de piétons. Les résultats de l'évaluation indiquent que cette méthode est performante face aux bruits des processus d'extraction de caractéristiques ; l'intégration de la connaissance passée et la variation d'échelle augmentent les performances de ce système de détection des piétons. Même dans un environnement routier encombré, une version quantifiée du système donne des résultats prometteurs.

Les évolutions incluent notamment des améliorations de la mise en œuvre Adaboost pour renforcer l'exactitude de la détection Bayésienne ; l'intégration du pistage multi-cible pour être en mesure de tester l'intégralité du système dans des environnements dynamiques avec le véhicule en déplacement.

Ce travail n'est actuellement pas terminé. L'idée de combiner la vraisemblance issue de la

détection d'un piéton dans les données LIDAR avec celle correspondant au score d'un classifieur image a été bien accueillie par la communauté (2 communications à des conférences internationales à comité de lecture). Il met en œuvre une version simplifiée du système de perception attentionnelle où une seule focalisation est effectuée. Il y ajoute un processus *bottom-up* : les piétons sont tout d'abord détectés par une analyse exhaustive d'une partie des données LIDAR correspondant à la zone d'analyse de la scène utile à l'application. Les hypothèses qui prennent ainsi naissance dans la scène sont confirmées ou infirmées par une analyse d'image focalisée dans les zones d'intérêt géométrique qui leur correspondent.

Le suivi d'objets se fait de manière naturelle à partir de l'état courant de l'analyse de la scène auquel est appliqué un processus d'évolution dynamique entre l'instant courant et la date d'acquisition nouvelles données en provenance des capteurs. Ce processus constitue donc une contribution originale au domaine de la reconnaissance/suivi multi-objet multi-capteur.

Perspectives

A courts termes

Apports de la communication entre entités constituant la scène

Parmi les objets dynamiques qui doivent être perçus, il est un cas particulier : celui des autres robots ou véhicules équipés d'un système de perception. Si de plus ces robots sont capables de communiquer entre eux (par exemple grâce à une liaison WiFi), ils peuvent s'échanger leur propre carte de l'environnement. Un robot dispose alors de sa propre carte et de celles qu'il a reçues. Ces cartes délocalisées représentent l'état de la scène du point de vue de chaque robot. Elle peuvent contenir des informations complémentaires et/ou redondantes. Leur fusion permet d'améliorer significativement la précision et la complétude de la carte d'un robot. Le système de perception qui s'appuie sur cette carte pour focaliser les observations voit ses performances augmenter.

Ce principe, actuellement développé dans la thèse de Nadir KARAM (que je co-encadre à 33%, soutenance prévue début 2008) est mis en oeuvre pour la localisation collective d'un groupe de robots communicants. Il apporte une contribution au problème important que pose pour la fusion d'information délocalisées, celui de la consanguinité des informations. Un premier robot reçoit les informations d'un second. Ces dernières peuvent avoir été obtenues par le second par fusion des informations précédemment émises par le premier. Chaque robot peut donc fusionner avec ses données locales des informations reçues construites à partir de données qu'il a lui même émises. Sans précaution ceci conduit à une instabilité dans l'estimation de l'état de la scène. La solution originale apportée à ce problème constitue la principale contribution de ce travail. Elle repose sur un principe simple : un robot ne communique jamais de données fusionnées. Il transmet uniquement l'information qu'il a construite à partir de son propre système de perception. Lorsqu'il communique, chaque robot dispose donc de plusieurs vues indépendantes de l'état de la scène. Il peut les fusionner sans risque de consanguinité. Il conserve l'état fusionné plus précis pour sa seule utilisation "personnelle" pour par exemple assurer son guidage automatique.

De plus, les données, qu'elles proviennent du système de perception local d'un robot ou qu'elles lui soient communiquées par d'autres, sont datées. Ceci permet de les prendre en compte de manière complètement asynchrone en fonction de leur disponibilité. Ce système gère donc naturellement les ruptures temporaires du flux d'information comme les pertes de communication WiFi ou l'impossibilité de percevoir telle ou telle partie de la scène. Ce dernier aspect constitue une autre contribution importante de cette méthode. Ce travail en cours de mise en œuvre expérimentale a fait l'objet de 3 communications dans des conférences internationales à comité de sélection.

Sur une plan plus général, il est tout à fait possible de considérer que le robot dispose, par le biais du canal de communication, d'une source supplémentaire d'informations extéroceptives que le système de perception attentionnelle, peut solliciter au même titre que n'importe quel capteur. En associant à cette information une fonction de focalisation, une fonction de coût, une utilité et une fonction de mise à jour, elle s'intègre de manière parfaitement rigoureuse au système de perception.

Apports de la programmation parallèle

La mise en oeuvre séquentielle actuelle du système de perception ne permet de suivre qu'une hypothèse de perception à la fois (à un moment donné, une seule branche de l'arbre est suivie). Il apparaît que la possibilité d'effectuer des calculs en parallèle conduit à émettre au même instant plusieurs hypothèses de perception. Chaque hypothèse peut générer sa propre chaîne d'hypothèses indépendamment les unes des autres.

L'idée est de construire une mise en oeuvre différente basée sur une parallélisation à trois niveaux algorithmiques :

1. Un parallélisme à bas niveau peut être mis en place pour chaque détecteur ce qui permet de réduire son temps de calcul.
2. Un parallélisme à moyen niveau qui consiste à appliquer en même temps l'ensemble des détecteurs d'une caractéristique dans la zone d'intérêt courante.
3. Un parallélisme à haut niveau permet la gestion multi hypothèses c'est-à-dire le parcours simultané de plusieurs branches de l'arbre des hypothèses. Dans la conception actuelle de l'algorithme, ceci signifie que plusieurs caractéristiques peuvent être observées en même temps.

Ce travail, confié à David GRIMBICHLER dans le cadre de la préparation d'une thèse que je co-encadre à 50% et qui a débuté en 2006, vise à étudier l'influence de l'approche parallèle sur la structure de l'algorithme de reconnaissance. Par exemple, les hypothèses doivent pouvoir « communiquer » entre elles c'est-à-dire que l'infirmation ou la confirmation de l'une peut avoir des conséquences sur les autres. De même, l'ensemble des détections déjà effectuées doit être accessible à toutes les hypothèses à venir pour éviter qu'un même détecteur soit appliqué plusieurs fois sur les mêmes données.

Il est communément admis qu'une architecture parallèle permet de raccourcir le temps de calcul. Le problème posé dans cette thèse peut se formuler par la question : "dans le cadre d'un système de perception pour un robot mobile, le parallélisme peut-il faire gagner autre chose que du temps ? "

A plus long termes : vers la perception cognitive

La cognition est définie comme "la génération de connaissances sur la base de la perception, du raisonnement, de l'apprentissage et des modèles *a priori*". Le système de perception proposé possède les caractéristiques d'un système cognitif.

Une perspective est d'intégrer plus systématiquement cet aspect dans le système. Ainsi dans l'application de localisation à partir d'une carte de triplets perceptifs il peut être intéressant

de construire automatiquement cette carte. Puisque les triplets peuvent reconnaître les amers qui leur correspondent, ils peuvent être appliqués dans le cas d'un processus *bottom-up* qui analyse exhaustivement les données capteurs pour en extraire des amers candidats avec une incertitude et une vraisemblance associée. La carte des triplets ainsi constituée sert bien sûr au guidage du système de perception attentionnel, tout en s'affinant simultanément. Il en résulte une cartographie d'amers labélisés (arbre, mûr, etc…, piétons, véhicules, etc …). Ceci peut constituer une alternative au SLAM (Simultaneous Localization And Mapping) et au SLAMMOT (Simultaneous Localization And Mapping with Mobile Object Tracking) pour laquelle les objets de la carte ne sont pas simplement des nuages de points 3D mais possédent un fort contenu sémantique.

Complétée en temps réel avec les objets dynamiques détectés dans l'environnement, cette carte constituerait un atout majeur pour instancier correctement et de manière cohérente les éléments clés de la scène constituant ainsi un véritable système de perception cognitif à même d'apprendre son environnement pour mieux le comprendre et ainsi interagir très finement avec lui.

Conclusion

Le système de perception attentionnel par focalisation progressive qui est le fil conducteur des applications présentées dans ce mémoire et accessoirement celui des travaux de recherche que je mène depuis une dizaine d'années au LASMEA, est un processus hautement actif où le système n'est pas le simple témoin des évolutions de son environnement, qu'il se contenterait de surveiller de manière systématique et en boucle ouverte sans réagir. Au contraire, il met en œuvre un contrôle poussé de l'acte de perception artificielle dans le but d'optimiser la prise de connaissance sur l'environnement.

Dans le contexte de la robotique mobile, ce système été conçu de manière globale et ne peut être considéré comme un enchaînement de briques élémentaires. Il est d'ailleurs revendiqué de manière forte en tant que système, ce qui signifie qu'il se veut en rupture avec la si rassurante approche modulaire. Convaincre de l'intérêt de ce type de système n'est pas chose facile. La communauté a ses habitudes pour ne pas dire ses effets de mode et il est vrai que pour une application donnée (par exemple la reconnaissance d'objet) ses résultats ne rivalisent pas (encore !) avec certaines approches très optimisées tendant vers un but unique.

Cependant, ces approches, de l'avis même de certains de leurs auteurs semblent en bout de course. L'énergie qu'il faut dépenser pour les faire encore progresser est sans doute démesurée par rapport au gain qui peut en résulter. Précurseur depuis longtemps, Olivier Faugeras l'a bien compris. Ainsi, après avoir pendant des années apporté des contributions immenses à la vision artificielle qui sont incarnées par le formalisme remarquable de la vision géométrique, a-t-il changé radicalement son fusil d'épaule en étudiant de manière conjointe la vision algorithmique et la vision humaine dans le but de s'inspirer de la seconde pour faire progresser la première [30].

D'autres auteurs beaucoup plus polémiques [29] (cités dans [46] p. 19) remarquent qu'après 40 ans d'active recherche internationale, la vision par ordinateur n'a rien produit de véritablement utile. . .

Nous croyons à cette approche système pour la perception en vue d'applications de robotique mobile. Notre ambition à terme est de proposer la **perception attentionnelle par focalisation progressive** comme un modèle [4] original de perception cognitif générique.

Le résultat principal du travail présenté réside dans la formalisation du processus de focalisation de l'attention selon un modèle descendant. Le contexte multisensoriel dans lequel il est parfois appliqué constitue également un aspect important de ce travail. Les applications de perception des objets dynamiques qui ont été présentées mettent en œuvre en complément un processus ascendant.

Des contributions reconnues ont été apportées dans chaque application présentée. La valori-

[4] Le terme paradigme est parfois utilisé.

sation expérimentale systématique, le plus souvent en temps réel embarqué dans des robots mobiles ou des véhicules démonstrateurs, qui a été menée dans chaque cas, témoigne de la volonté de ne pas se contenter de la solution de facilité que constitue la simulation ou la seule démonstration sur table fût-elle sur des données réelles. Elle suffit me semble-t-il à crédibiliser l'approche au moins autant qu'un grand nombre de publications. Les industriels avec qui nous avons pu collaborer ne s'y sont pas trompés puisqu'ils nous ont eux mêmes sollicité dans de nombreux cas.

De manière plus générale, il faut considérer que la recherche en Sciences pour l'Ingénieur et en Sciences de l'Information et de la Communication n'existe pas si elle n'est pas appliquée et si elle ne se pose pas la question de son utilité sociale. Autant les sciences fondamentales peuvent se satisfaire du seul but de comprendre notre univers, autant les acteurs de la "*science des systèmes créés par l'homme*" ont le devoir le produire des résultats uties. En d'autres termes plus que d'autres, ils doivent s'interroger sur ce qu'ils produisent.

Bibliographie

[1] *Le petit Larousse Illustré*. Larousse, 2004.

[2] Broggi A. An image reorganization procedure for automotive road following systems. In *Int. Conf. on Image Processing*, Oct. 22–25 1995.

[3] R. Aufrere, R. Chapuis, and F. Chausse. A model-driven approach for real-time road recognition. *International Journal Machine Vision Applications*, 13(2) :95–107, November 2001.

[4] I. Autio and K.T. Lindgren. Attention-driven parts-based object detection, 2004.

[5] R. Bajcsy. Active perception. In *PROC IEEE*, volume 76, pages 996–1005, 1988.

[6] D. H. Ballard and C. M. Brown. Principles of animate vision. *CVGIP*, 56(1) :3–21, July 1992.

[7] B. Barshan and H.F. Durrant. *An inertial navigation system for a mobile robot*. In *IFAC International Worshop On Intelligent Autonomous Vehicles*, Southampton UK, 1995.

[8] M. Bertozzi, A. Broggi, R. Chapuis, F. Chausse, A. Fascioli, and A. Tibaldi. Shape-based pedestrian detection and localization. *In Procs. IEEE Intl. Conf. on Intelligent Transportation Systems 2003*, pages 328–333, 2003.

[9] P. Bonnifait. Localisation précise en position et attitude des robots mobiles d'extérieur. PhD thesis, Ecole centrale de Nantes, France, 1997.

[10] GA. Borges and MJ. Aldon. Robustified estimation algorithms for mobile robot localization based on geometrical environment maps. *Robotics and Autonomous Systems*, 45(3-4) :131–159, December 2003.

[11] S. Botton, F. Duquenne, Y. Egels, M. Even, and P. Willis. *GPS localisation et navigation*. Hermes, 1998.

[12] D. Bouvet. Contribution à la localisation d'engins de chantiers routiers. PhD thesis, Ecole centrale de Nantes, France, 2000.

[13] T.M. Breuel. Model based recognition using pruned correspondence search. In *CVPR91*, pages 257–262, 1991.

[14] A. Broggi, M. Bertozzi, A. Fascoli, and M. Sechi. Shape-based pedestrian detection. *In Proc. IEEE Intelligent Vehicle Symposium 2000*, pages 215–220, 2000.

[15] A. Brunnert. Machine guidance with laser and gps. In *Conference on Crop Harvesting and Processing*, Kentucky (USA), February 2003.

[16] M.C. Burl, M. Weber, and P. Perona. A probabilistic approach to object recognition using local photometry and global geometry. In *Proc. 5th European Conf. Comp. Vision*, pages 628–641, 1998.

[17] Koch C. and Ullman S. Shift in selective visual attention : towards the underlying neural circuitry. *Human Neurobiology*, 4 :229–227, 1985.

[18] Jensfelt Christensen. Active global localization for mobile robots using multiple hypothese tracking. In *IEEE Transactions on Robotics and Automation*, 2002.

[19] M. Chung, L. Ojeda, and J. Borenstein. Sensor fusion for mobile robot dead-reckoning with a precision-calibrated fiber optic gyroscope. In *IEEE International Conference on Robotics and Automation*, pages 3588–3593, Korea, May 2001.

[20] Dissanayake G. Newman P. Clark, S. and H. Durrant-Whyte. A solution to simultaneous localisation and map building (slam) problem. *IEEE Journal of Robotics and Automation*, 17(3), 2001.

[21] Tessier Cédric. *Système de localisation basé sur une stratégie de perception cognitive appliquée à la navigation autonome d'un robot mobile*. Novembre 2007.

[22] Dickmanns E. D. and Mysliewetz B. D. Recursive 3d road and relative ego-state recognition. *Trans. on Pattern Analysis and Machine Intelligence*, 14(2) :199–213, 1992.

[23] Pommerleau D.A. Ralph : Rapidly adapting lateral position handler. In *IEEE Symposium on Intelligent Vehicles*, Detroit, Michigan (USA), 1995.

[24] AJ. Davison and DW. Murray. Simultaneous localization and map-building using active vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7) :865–880, July 2002.

[25] G. Deco and B. Sch¸rmann. A hierarchical neural system with attentional top-down enhancement of the spatial resolution for object recognition. *Vision Research*, 40 :2845–2859, 2000.

[26] S.J. Dickinson, H.I. Christensen, J.K. Tsotsos, and G. Olofsson. Active object recognition integrating attention and viewpoint control. *Computer vision and image understanding*, 67(3) :239–260, September, 1997.

[27] B. Draper, S. Buluswar, A. Hanson, and E. Riseman. Information acquisition and fusion in the mobile perception laboratory. In *SPIE Sensor Fusion VI*, Boston, MA, 1993.

[28] Cyril Drocourt. *Localisation et modélisation de l'environnement d'un robot mobile par coopération de deux capteurs omnidirectionnels*. PhD thesis, Université de Technologie de Compiègne - Centre de Robotique d'Electronique et d'Automatique, février 2002.

[29] M. Duff. Forum. In *IAPR Newsletter*, volume 21, 1999.

[30] T. Vieville et O. Faugeras. Perception visuelle en robotique : Profiter de la biologie pour faire des systèmes adaptatifs. Technical report, INRIA - Projet Odyssée, Octobre 2001.

[31] P. Felzenszwalb and D. Huttenlocher. Pictorial structures for object recognition, 2003.

[32] D. Fox, J. Hightower, L. Liao, D. Schulz, and G. Borriello. Bayesian filtering for location estimation. *IEEE Pervasive Computing*, pages 10–19, 2003.

[33] Dieter Fox, Wolfram Burgard, and Sebastian Thrun. Active markov localization for mobile robots. *Robotics and Autonomous Systems*, 25 :195–207, 1998.

[34] Udo Frese. A discussion of simultaneous localization and mapping. *Autonomous Robots*, 20 :25–42, 2006.

[35] Yoav Freund and Robert E.Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, (55), 1997.

[36] S. Frintrop and E. Rome. Simulating visual attention for object recognition.

[37] K.C. Fuerstenberg and U. Lage. Pedestrian detection and classification by laserscanners. *9th EAEC International Congress*, 2003.

[38] T. Gandhi and M. Trived. Pedestrian collision avoidance systems : A survey of computer vision based recent studies. *In Proc. of the IEEE Intelligent Transportation Systems Conference*, pages 976–981, 2006.

[39] D.M. Gavrila. Sensor-based pedestrian protection. *IEEE Intelligent Systems*, 16 :77–81, 2001.

[40] W.E.L. Grimson and T. Lozano-Perez. Model-based recognition and localization from sparse range or tactile data. In *RCV87*, pages 382–414, 1987.

[41] F. Guichard and J.P. Tarel. Curve finder combining perceptual grouping and a kalman like fitting. In *ICCV99*, pages 1003–1008, 1999.

[42] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE transactions on pattern analysis and machine intelligence*, 20(11) :1254–1259, November 1998.

[43] B. Everett J. Borenstein and L. Feng. *"Where I am ?" Sensors and methods for mobile robot positionning*. In *Technical Report*. University of Michigan, 1996.

[44] Andrew J.Davison. Active search for real-time vision. *ICCV*, 2005.

[45] Laneurit Jean. *Fusion d'informations pour la localisation d'un véhicule*. Juillet 2006.

[46] J.M. Jolion (sous la direction de). *Les systèmes de vision*. Hermes Science Publications, 2001.

[47] Jose A. Castellanos Kai O. Arras and Roland Siegwart. Feature-based multi-hypothesis localization and tracking for mobile robots using geometric constraints. In *International Conference on Robotics and Automation*, Washington DC, USA, May 2002.

[48] Chen K.H. and Tsai W.H. Vision-based autonomous land vehicle guidance in outdoor road environments using combined line and road following techniques. *Journal of Robotics Systems*, 14(10) :711–728, 1997.

[49] C. Kreucher and S. Lakshmanan. A frequency domain approach to lane detection in roadway images. In *International Conference on Image Processing*, pages 31–35, KOBE, October 24–28 1999.

[50] Jean Laneurit, Roland Chapuis, and Frédéric Chausse. Accurate vehicle positioning on a numerical map. *International Journal of Control, Automation, and Systems*, 3(1) :15–31, March 2005.

[51] T.K. Leung, M.C. Burl, and P. Perona. Finding faces in cluttered scenes using random labelled graph matching. In *In fifth Intl. Conf. on Computer Vision.*, pages 637–644, 1995.

[52] Joseph Machrouh and Philippe Tarroux. Attentional mechanisms for interactive image exploration. *EURASIP Journal on Applied Signal Processing*, 2005(14) :2391–2396, 2005.

[53] M.DI Marco, A. Garulli, A. Giannitrapani, and A. Vicino. A set theoretic approach to dynamic robot localization and mapping. *Autonomous Robots*, 16 :23–47, 2004.

[54] D. Marr. *Vision*. W.H.Freeman and Company, 1982.

[55] A. Mendes and U. Numes. Situation-based multi-target detection and tracking with laserscanner in outdoor semi-structured environment. *In Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2004.

[56] A. Mohan, C. Papageorgiou, and T. Poggio. Example-based object detection in images by components. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(4) :349–361, 2001.

[57] G. Monteiro, C. Premebida, P. Peixoto, and U. Numes. Tracking and classification of dynamic obstacles using laser range finder and vision. *In Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006.

[58] M. Murphy, A. Torralba, and W. Freeman. Using the forest to see the trees : a graphical model relating features, objects, and scenes. volume 16. MIT Press, 2003.

[59] B. Nabbe and M. Hebert. Where and when to look. In *IROS*, October 2003.

[60] J. Neira and Juan D. Tardos. Data association in stochastic mapping using the joint compatibility test. *IEEE Transactions on Robotics and Automation*, 17(6) :890 ?897, 2001.

[61] Paul Newman and Kin Leong Ho. Slam-loop closing with visually salient features. In *International Conference on Robotics and Automation*, Barcelona, Spain, April 2005.

[62] Juan Nieto, Jose Guivant, Eduardo Nebot, and Sebastian Thrun. Real time data association for fastslam. In *International Conference on Robotics & Automation*, septembre 2003.

[63] Trujillo Noël. *Stratégie de perception pour la compréhension de scènes par une approche focalisante, application à la reconnaissance d'objets*. Décembre 2007.

[64] M. O'Connor, G. Elkaim, T. Bell, and B. Parkinson. Automatic steering of a farm vehicle using gps. In 3^{nd} *International Conference On Precision Agriculture*, pages 767–777, June 1996.

[65] C.P. Papageorgiou, M. Oren, and T. Poggio. A general framework for object detection. In *ICCV '98 : Proceedings of the Sixth International Conference on Computer Vision*, page 555, Washington, DC, USA, 1998. IEEE Computer Society.

[66] P.Viola, M. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. *In Proc. of the International Conference on Computer Vision (ICCV)*, 2003.

[67] P.Viola and M.Jones. Rapid object detection using a boosted cascade of simple features. *In IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

[68] Chapuis R. *Suivi de primitives image, application ‡ la conduite automatique sur route*. PhD thesis, Univ. B. Pascal, Clermont-Ferrand France, jan. 1991.

[69] O. Ramstrˆm and H.I. Christensen. Object detection using background context. In *Procedings of the 17th International Conference on Pattern Recognition (ICPR'04). IEEE*. Computer Society IEEE, 2004.

[70] J. Ried and D. Niebuhr. Driverless tractors. In *Ressource*, volume 8 of *7-8*, 2001.

[71] Chapuis Roland. *Localisation et suivi de route pour l'aide à la conduite l'aide à la conduite*. Décembre 2000.

[72] Aufrère Romuald. *Reconnaissance et suivi de route par vision artificielle, application à l'aide à la conduite*. Juin 2001.

[73] H. Schneiderman and T. Kanade. Object detection using the statistics of parts. *International Journal of Computer Vision*, 56(3) :151–177, 2004.

[74] T. Serre, L. Wolf, and T. Poggio. A new biologically motivated framework for robust object recognition. Technical report, 2004.

[75] Y. Sun and R. Fisher. Object-based visual attention for computer vision. *Informatics research report EDI-INF-RR-0213*, June, 2004.

[76] M. Szarvas, U. Sakai, and J. Ogata. Real-time pedestrian detection using lidar and convolutional neural networks. *In IEEE Intelligent Vehicles Symposium 2006*, 2006.

[77] B. Takacs and H. Wechsler. A dynamical and multiresolution model of visual attention and its application to facial landmark detection. *Computer Vision and Image Understanding*, 70(1) :63–73, April, 1998.

[78] R. Talluri and J.K. Aggarwal. *Image/Map Correspondence for Mobile Robot Self-Location Using Computer Graphics*. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1993.

[79] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. *Probabilistic Robotics*. Intelligent Robotics and Autonomous Agents. MIT Press, September 2005.

[80] J.K. Tsotsos, S.M. Culhane, W. Yan Key Wai, Y. Lai, N. Davis, and F. Nuflo. Modeling visual attention via selective tuning. *Artificial intelligence*, 78 :507–545, 1995.

[81] J. Vaganay. Conception d'un système multisensoriel de localisation dynamique 3D pour robot mobile. PhD thesis, University Montpellier II, France, 1993.

[82] J. Vaganay, J.G. Belligham, and J. Leonard. *Comparison of fix computation and filtering for autonomous acoustic navigation*. In *International Journal of Systems Science*, 1998.

[83] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Conference on computer vision and pattern recognition*, volume 1, pages 511–518.

[84] B.A. w and A. Lionelle. Evaluation of selective attention under similarity transformations. *Computer vision and image understanding.*, 100 :152–171, 10 August 2005.

[85] D. Walther, L. Itti, M. Riesenhuber, T. Poggio, and C. Koch. Attentional selection for object recognition- a gentle way. In *Second IEEE International Workshop, BMCV*, pages 472–479, 2002.

[86] C.-C. Wang, C. Thorpe, and S. Thrun. Online simultaneous localization and mapping with detection and tracking of moving objects : Theory and results from a ground vehicle in crowded urban areas. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Taipei, Taiwan, 2003.

[87] H. Wang, T. Ishimatsu, and J.T. Mian. *Self-Location for an electric wheel-chair*. In *Japan Society of Mechanical Engineers International Journal*, 1997.

[88] Y. Aloimonos I. Weiss and A. Bandopadhay. Active vision. In *Int'l J. Computer Vision*, volume 1, pages 333–356, Jan. 1987.

[89] L. Zhao and C. Thorpe. Stereo- and neural network-based pedestrian detection. *IEEE Transaction on Intelligent Transportation Systems*, 2000.